# БИОМОРФНЫЙ НЕЙРОПРОЦЕССОР НА ОСНОВЕ НАНОРАЗМЕРНОГО КОМБИНИРОВАННОГО МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА

**А.Д. Писарев, С.Ю. Удовиченко**

**А.Д. Писарев, С.Ю. Удовиченко**

# Биоморфный нейропроцессор на основе наноразмерного комбинированного мемристорно-диодного кроссбара

**УДК 004.032.26**
**ББК 16.632**
**П34**

**П34 Писарев А.Д., Удовиченко С.Ю.**
**Биоморфный нейропроцессор на основе наноразмерного комбинированного мемристорно-диодного кроссбара**

Предложена концепция биоморфного нейропроцессора, реализующего аппаратную импульсную нейросеть для традиционных задач обработки информации, а также для воспроизведения работы кортикальной колонки мозга или её фрагмента. Аппаратная нейросеть построена на основе оригинальных биоморфных программной и электрической моделей нейрона. Представлены электрические схемы, топология и нанотехнология изготовления основных узлов аппаратной части нейропроцессора: запоминающей и логической матриц, входного и выходного устройств, построенных на основе комбинированного мемристорно-диодного кроссбара и обладающих высокими интеграцией элементов и энергоэффективностью по сравнению с известными нейропроцессорами и отдельными матрицами. Приведены результаты SPICE-моделирования и аппаратного тестирования процессов обработки сигналов в режимах: сложения выходных импульсов нейронов в запоминающей матрице; их маршрутизации на синапсы других нейронов в логической матрице, скалярного умножения матрицы чисел на вектор, а также ассоциативного самообучения. Впервые продемонстрирована генерация новой ассоциации (нового знания) в изготовленном мемристорно-диодном кроссбаре в отличие от ассоциативного самообучения в существующих аппаратных нейросетях с синапсами на базе дискретных мемристоров.

The concept of a biomorphic neuroprocessor that implements a hardware pulsed neural network for traditional information processing tasks, as well as for reproducing the operation of the brain cortical column or its fragment, is proposed. The hardware neural network is based on the original biomorphic software and electrical models of the neuron. Electrical circuits, topology, and fabrication nanotechnology of main nodes of the neuroprocessor are presented. The main nodes are memory and logic matrices, input and output devices and are built on the basis of a composite memristor-diode crossbar and have high element integration and energy efficiency compared to known neuroprocessors and separate matrices. The results of SPICE simulation and hardware examination of signal processing routines in the following modes are presented. The simulated modes are summation of output pulses of neurons in a memory matrix, their routing to synapses of other neurons by the logic matrix, matrix-vector dot product, and associative self-learning. The generation of a new association (new knowledge) in a manufactured memristor-diode crossbar is demonstrated for the first time, as opposed to associative self-learning in existing hardware neural networks with synapses based on discrete memristors.

**УДК 004.032.26**
**ББК 16.632**

**ISBN 978-5-94836-635-7**

# Оглавление

# Предисловие

Настоящая монография написана по материалам научных статей авторов, опубликованных в высокорейтинговых журналах, рекомендованных ВАК РФ и входящих в базы Web of Science и Scopus. Инновационная разработка биоморфного нейропроцессора проводилась в научной группе, возглавляемой доктором физико-математических наук, профессором кафедры прикладной и технической физики, руководителем Научно-образовательного центра «Нанотехнологии» Тюменского государственного университета Сергеем Юрьевичем Удовиченко.

В состав группы входили: кандидат технических наук, доцент кафедры прикладной и технической физики, зав. лабораторией пучково-плазменных технологий НОЦ «Нанотехнологии» Александр Дмитриевич Писарев; кандидат технических наук, зав. лабораторией электронной и зондовой микроскопии НОЦ «Нанотехнологии» Андрей Николаевич Бобылев, а также аспиранты кафедры прикладной и технической физики Александр Николаевич Бусыгин, Абдулла Хойдар Ибрагим и Алексей Александрович Губин.

Инновационная разработка биоморфного нейропроцессора выполнена в рамках Национальной технологической инициативы NeuroNet, поддержанной 09.06.2015 г. на заседании Президиума Совета при Президенте Российской Федерации по модернизации экономики и инновационному развитию России. Изготовление наноматериалов и электронных устройств, предназначенных для аппаратной реализации нейропроцессора, проводилось на инновационном оборудовании — Нанотехнологическом комплексе NT-MDT «НаноФаб-100», приобретенном в рамках Федеральной целевой программы «Развитие инфраструктуры наноиндустрии в России с 2007 по 2011 г.».

Фундаментальные исследования и аппаратное тестирование электронных устройств нейропроцессора осуществлялись в рамках трех проектов, поддержанных грантами РФФИ: № 19-07-00272 «Электрофизические свойства комбинированного мемристорного-диодного кроссбара — нового компонента наноэлектроники, предназначенного для изготовления запоминающей и логической матриц нейропроцессора», № 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» и № 20-37-90003 «Моделирование физических процессов в мемристорно-диодных кроссбарах входного и выходного блоков нейропроцессора».

В 2019 г. за инновационную разработку биоморфного нейропроцессора С.Ю. Удовиченко получил диплом лауреата Тюменского областного конкурса «Лидер в научно-инновационной деятельности» в номинации «Ученый года в инновациях».

# Введение

## ОТЛИЧИЕ БИОМОРФНОГО НЕЙРОПРОЦЕССОРА, СПОСОБНОГО ВОСПРОИЗВОДИТЬ РАБОТУ КОРТИКАЛЬНОЙ КОЛОНКИ, ОТ ПРИМЕНЯЕМЫХ В IT НЕЙРОПРОЦЕССОРОВ

Под нейропроцессором подразумевается автономное аппаратное средство, предназначенное для решения нейросетевых задач.

В настоящее время развиваются два подхода к моделированию нейрона: информационный и биологический. В информационном подходе, в котором используется модель формального нейрона, уровень упрощения слишком высок: дендриты и аксоны заменены на связи с единственной характеристикой — весом связи. Обучение нейросети заключается в подборе весов по определенным правилам [1].

Современная биологическая модель нейрона [2] является упрощенной по сравнению с моделью Hodgkin, Huxley [3], в которой изменение потенциала во времени на мембране нейрона описывается двумя дифференциальными уравнениями первого порядка с восемью параметрами. В базовой модели [3] использовалось пять дифференциальных уравнений первого порядка с тридцатью одним параметром. Переход к упрощенной модели был обоснован в работах [4; 5], где показано, что часть параметров слабо влияет на поведение нейрона.

Попытка аналогового решения системы дифференциальных уравнений на специализированном устройстве с использованием операционных усилителей приводит к большому потреблению энергии и низкой степени интеграции элементов [6]. Для обработки информации в сверхбольшой нейросети, предназначенной для нейропроцессора, необходимо построить модель нейрона максимально упрощенную (с точки зрения времени расчета), но без существенной потери точности. Этого можно достичь путем замены численного решения дифференциальных уравнений, описывающих изменение потенциала на мембране нейрона, на явные передаточные функции.

В работе [7] нами представлена оригинальная биоморфная модель нейрона, состоящая из отдельных функциональных частей — дендритов, сомы и аксона. Созданы алгоритмы расчета прохождения сигнала через каждую функциональную часть нейрона и вместо системы дифференциальных уравнений построены соответствующие передаточные функции для потенциала мембраны в виде рекуррентных соотношений. Такой подход существенно сокращает время расчетов в нейросети и позволяет реализовывать

любые соединения между отдельными частями разных нейронов, что придает большую гибкость архитектуре биоморфной нейросети.

Показано, что кодирование передаваемой информации импульсами, подобными биологическим, позволяет использовать мемристоры для расчета рекуррентных формул, описывающих изменение количества рецепторов на мембране дендрита. Разработанная биоморфная модель нейрона, сформулированные концептуальные принципы построения нейросети на ее основе, а также замена синапсов в аппаратной нейросети на мемристоры позволяют построить сверхбольшую импульсную нейросеть, моделирующую работу отдельной кортикальной колонки на автономном аппаратном средстве.

Использование мемристоров в качестве синаптических связей [8] в биоморфной модели нейрона потенциально может увеличить быстродействие и снизить потребление энергии электронного устройства. Это достигается за счет замены программного расчета прохождения сигнала и характеристик синаптической связи, выполняемого в транзисторной электрической схеме, на прямое прохождение сигнала через мемристор, свойства которого изменяются аналогично оригинальной синаптической связи.

Импульсные нейронные сети являются более биологически правдоподобными, используют меньшее количество нейронов, но требуют большего объема программных вычислений. В импульсных нейросетях реализуется биологически подобный механизм самообучения, который сложно реализовать в традиционных сетях с точечными нейронами [9]. Импульсные нейронные сети превосходят нейросети на точечных нейронах в точности и вычислительной мощности и лучше приспособлены для аппаратной реализации из-за работы по принципу “integrate-and-fire” [10].

Исследования по созданию автономного аппаратного средства, предназначенного для программно-аппаратной реализации нейросети, в последнее время проводятся достаточно интенсивно. Но до сих пор, кроме настоящей разработки авторов [11], не представлено автономное аппаратное средство, реализующее самообучающуюся импульсную нейронную сеть с программируемыми синаптическими связями на основе мемристоров.

Нейропроцессор IBM TrueNorth [12], построенный на транзисторах по КМОП-технологии, благодаря многоядерной архитектуре обеспечивает достаточную производительность для моделирования кортикальной колонки мозга. Архитектура аппаратного средства для имитации нейронной сети, представленная в патенте [13], является последней разработкой этой фирмы, в которой синаптические связи по-прежнему реализуются с помощью транзисторов. Разработчики этих аппаратных средств планируют применить мемристоры в качестве синапсов, что обеспечит значительное сокращение числа транзисторов, высокую интеграцию элементов и энергоэффективность устройства. Такое сокращение числа транзисторов возможно благодаря реализации в мемристоре множества резистивных состояний. В этом нейропроцессоре не заложена способность к самообучению

в процессе обработки информации. Весовые коэффициенты синапсов переносятся в нейропроцессор после предварительного программного обучения нейросети на компьютере.

Последними образцами импульсных нейропроцессоров, изготовленными так же, как и TrueNorth на транзисторах по КМОП-технологии, являются Brainchip Akida [14] и Intel Loihi [15]. В этих нейропроцессорах уже реализованы механизмы самообучения импульсных нейросетей. Концепция нейроморфных устройств, являющихся комбинацией КМОП-логики и мемристоров, впервые предложена в [16; 17] и реализована в виде автономного устройства в [18]. Существующие аппаратные средства в виде запоминающих [19–22] и логических матриц [23–26] на мемристорах выполняют узкоспециальные функции. Наиболее продвинутый массив мемристорных ячеек, впервые предложенный Hewlett-Packard как аппаратное средство для выполнения дискретного косинусного преобразования при обработке сигналов (сжатие изображений и сверточная фильтрация) [26], был использован в дальнейшем как массив синапсов аппаратной нейросети, которая обучается с учителем [21]. Кроме того, эту матрицу можно использовать для выполнения логических операций при подаче входных логических сигналов на затворы транзисторов. Это универсальное аппаратное средство может выполнять отдельные функции предлагаемого в настоящей работе биоморфного нейропроцессора и обладает низкой интеграцией элементов и высоким энергопотреблением из-за применения операционных усилителей.

В работе [11] нами 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 аудиоинформации, а также для воспроизведения работы кортикальной колонки мозга или ее фрагмента. В качестве ключевых узлов аппаратной части нейропроцессора используются сверхбольшие запоминающая [27] и логическая [11] матрицы, представляющие собой массив синапсов и задающие вес и маршрут связи между нейронами соответственно. Указанные матрицы являются сверхбольшими потому, что каждый нейрон в сети может обладать большим количеством синаптических связей.

В биоморфном нейропроцессоре продемонстрированы высокие интеграция элементов и энергоэффективность по сравнению с известными нейропроцессорами и отдельными матрицами. Такая эффективность достигнута за счет применения смешанных аналогово-цифровых вычислений, в том числе с помощью мемристоров, интегрированных в мемристорно-диодные кроссбары.

Приведены результаты SPICE-моделирования и аппаратного тестирования процессов обработки сигналов в режимах: сложения выходных импульсов нейронов в запоминающей матрице; их маршрутизации на синапсы других нейронов в логической матрице, умножения матрицы чисел на вектор, которое применяется в обоих матрицах, и ассоциативного

самообучения. При сравнении экспериментальных результатов и SPICE-моделирования процессов обработки сигналов в мемристорно-диодном кроссбаре определен разброс параметров мемристоров, при котором наблюдается устойчивая работа кроссбара [28].

Ассоциативное самообучение и формирование новой ассоциации в нейросети с мемристорными синапсами по правилу Хебба впервые предложено в [29]. Эта идея развита и аппаратно реализована на дискретных мемристорах в работах [30–34]. 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.

В процессе разработки электрической схемы нейропроцессора [11] проведена модернизация биоморфной электрической модели нейрона Ходжкина–Хаксли, которая позволила реализовать ассоциативное самообучение в запоминающей матрице и условное самообучение нейронной сети в случае отсутствия запоминающей матрицы в схеме нейропроцессора.

Впервые продемонстрирована генерация новой ассоциации (нового знания) в изготовленном мемристорно-диодном кроссбаре в отличие от ассоциативного самообучения в существующих аппаратных нейросетях с синапсами на базе дискретных мемристоров.

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 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. По-существу представленный нейропроцессор является прототипом компьютеров нового поколения, являющихся носителями искусственного интеллекта.

Аналогов разработанного биоморфного нейропроцессора нет. Уникальность биоморфного нейропроцессора состоит в том, что он построен на основе модернизированной электрической биоморфной модели нейрона и является биоморфным еще и с точки зрения выполнения функций биоморфной нейросети, созданной на основе оригинальной программной биоморфной модели нейрона. Кроме задачи генерации нового знания, относящейся к проблеме создания систем искусственного интеллекта, на таком специализированном аппаратном средстве могут быть решены технические задачи — увеличение быстродействия и энергоэффективности

расчетов по сравнению с существующими сегодня вычислительными средствами (персональные компьютеры, серверы и суперкомпьютеры) — за счет применения смешанных аналогово-цифровых вычислений, в том числе с помощью новых элементов электроники — мемристоров, интегрированных в комбинированные мемристорно-диодные кроссбары.

Разработку нейропроцессоров на основе мемристоров и мемристорных кроссбаров тормозит то обстоятельство, что применяемые твердотельные мемристоры на оксидах переходных металлов пока имеют невысокую стабильность и воспроизводимость электрических характеристик. Однако предложенный биоморфный нейропроцессор не подвержен этому обстоятельству. Распределенный характер биоморфной нейросети снижает требования к воспроизводимости и стабильности характеристик мемристоров. Кроме этого, в аппаратную реализацию нейросети можно добавить электрическую схему, которая будет имитировать работу астроцита (один из видов глиальных клеток мозга) путем увеличения проводимости оставшихся мемристорных синапсов при обнаружении поврежденного [7; 35; 36]. Это повышает отказоустойчивость схемы нейропроцессора и дополнительно увеличивает биоморфность нейросети. Еще одна возможность коррекции поврежденного мемристорного синапса состоит в увеличении возбудимости пресинаптического нейрона путем снижения порога активации или увеличения весов его возбуждающих синапсов.

На фоне сверхбыстрой (порядка микросекунд) обработки информации программно-аппаратными нейросетями для ее распознавания и классификации, которая применяется сегодня в информационных технологиях (IT), в предложенной биоподобной системе формируются медленные процессы. Изменение электрического потенциала при нейронном возбуждении в точках нервного волокна происходит в виде спайкового импульса величиной десятки милливольт и с характерным временем 1 мс. Задержки сотни миллисекунд нужны для формирования паттернов спайков, которые предположительно используются в биологических системах как вариант кодирования информации для ее эффективного выделения в сравнительных операциях.

Развитие автономного ассоциативного процессора, являющегося носителем искусственного интеллекта, по мнению авторов монографии, будет проходить по биоподобному сценарию.

Разработанная и протестированная аппаратная биоморфная нейросеть нейропроцессора потенциально способна воспроизводить работу кортикальной колонки мозга и 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 и который обеспечит осмысливание полученных новых ассоциаций при совершенствовании его оригинальной биоморфной нейросети [7] и, следовательно, переход от слабого (*narrow*) к сильному (*general*) искусственному интеллекту.

Список литературы


1. *Rosenblatt F.* The perceptron: A probabilistic model for information storage and organization in the brain // Psychological Review. 1958. Vol. 65. Pp. 386–408.
2. *Brette R.* Philosophy of the spike: Rate-based vs. spike-based theories of the brain // Frontiers in Systems Neuroscience. 2015. Vol. 9. P. 151.
3. *Hodgkin A.L.*, *Huxley A.F.* A quantitative description of membrane current and its application to conduction and excitation in nerve // Journal of Physiology.1952. Vol. 117. No. 4. Pp. 500–544.
4. *Goldman M.S.*, *Golowasch J.*, *Marder E.*, *Abbott L.F.* Global structure, robustness, and modulation of neuronal models // Journal of Neuroscience. 2001. Vol. 21. Pp. 5229–5238.
5. *Prinz A.A.*, *Billimoria C.P.*, *Marder E.* Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons // Journal of Neurophysiology. 2003. Vol. 90. Pp. 3998–4015.
6. *Millner S.*, *Grubl A.*, *Meier K.* et al. A VLSI Implementation of the adaptive exponential integrate-and-fire neuron model // Proceedings of the 23rd International Conference on Neural Information Processing Systems. 2010. Pp. 1642–1650.
7. *Filippov V.A.*, *Bobylev A.N.*, *Busygin A.N.* et al. Neural biomorphic model and neural network construction principle for corticomorphic coprocessor based on memristor elements // Neural Computing and Applications. 2020. Vol. 32. Pp. 2471–2485.
8. *Bobylev A.N.*, *Udovichenko S.Yu.* The electrical properties of memristor devices $TiN/Ti_xAl_{1-x}O_y/TiN$ produced by magnetron sputtering // Russian Microelectronics. 2016. Vol. 45. No. 6. Pp. 396–401.
9. *Khacef L.*, *Abderrahmane N.*, *Miramond B.* Confronting machine-learning with neuroscience for neuromorphic architectures design // 2018 International Joint Conference on Neural Networks (IJCNN). 2018. P. 8489241.
10. *Lobo J.L.*, *Ser J.D.*, *Bifet A.*, *Kasabov N.* Spiking Neural Networks and online learning: An overview and perspectives // Neural Networks. 2020. Vol. 121. Pp. 88–100.
11. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor — diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
12. *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. Vol. 345. Pp. 668–672.
13. *Rivera R.A.-I.*, *Arthur J.V.*, *Cassidy A.S.* et al. Hardware architecture for simulating a neural network of neurons // 2019. US Patent 2019 197 394.
14. *Van Der Made P.A.J.*, *Viejo A.*, *Mankar A.S.*, *Viejo M.* Neural Processor based accelerator system and method // 2017. US Patent 0024644.
15. *Davies M.*, *Srinivasa N.*, *Lin T.-H.* et al. Loihi: A neuromorphic manycore processor with on-chip learning // IEEE Micro. 2018. Vol. 38. No. 1. Pp. 82–99.
16. *Prezioso M.*, *Merrikh-Bayat F.*, *Hoskins B.D.* et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors // Nature. 2015. Vol. 521. Pp. 61–64.
17. *Kim K.-H.*, *Gaba S.*, *Wheeler D.* et al. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications // Nano Letters. 2012. Vol. 12. Pp. 389–395.
18. *Bobylev A.N.*, *Busygin A.N.*, *Pisarev A.D.* et al. Neuromorphic coprocessor prototype based on mixed metal oxide memristors // International Journal of Nanotechnology. 2017. Vol. 14. No. 7/8. Pp. 698–704.

19. *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. Pp. 125–130.
20. *Yao P.*, *Wu H.*, *Gao B.* et al. Online training on RRAM based neuromorphic network: Experimental demonstration and operation scheme optimization // 2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM). 2017. Pp. 182–183.
21. *Li C.*, *Belkin D.*, *Li Y.* et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks // Nature Communications. 2018. Vol. 9. P. 2385.
22. *Ielmini D.* Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks // Microelectronic Engineering. 2018. Vol. 190. Pp. 44–53.
23. *Levy Y.*, *Bruck J.*, *Cassuto Y.* et al. Logic operations in memory using a memristive Akers array // Microelectronics Journal. 2014. Vol. 45. Pp. 1429–1437.
24. *Li C.*, *Hu M.*, *Li Y.* et al. Analogue signal and image processing with large memristor crossbars // Nature electronics. 2018. Vol. 1. No. 1. Pp. 52–59.
25. *Teimoori M.*, *Amirsoleimani A.*, *Ahmadi A.*, *Ahmadi M.* A 2M1M crossbar architecture: Memory // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2018. Vol. 26. No. 12. Pp. 2608–2618.
26. *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. Vol. 62. No. 5. Pp. 1402–1411.
27. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Yu.*, *Maevsky O.V.* 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor // Microelectronic Engineering.2018. Vol. 198. Pp. 1–7.
28. *Pisarev A.*, *Busygin A.*, *Bobylev A.* et al. 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. Vol. 236. P. 111471.
29. *Pershin Y.V.*, *Ventra M.D.* Experimental demonstration of associative memory with memristive neural networks // Neural Networks. 2010. Vol. 23. No. 7. Pp. 881–886.
39. *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. 2018. Vol. 65. No. 7. Pp. 2210–2220.
31. *Yang L.*, *Zeng Z.*, *Huang Y.*, *Wen S.* Memristor-based circuit implementations of recogni-tion network and recall network with forgetting stages // IEEE Transactions on Cognitive and Developmental Systems. 2018. Vol. 10. No. 4. Pp. 1133.
32. *Wang Z.*, *Rao M.*, *Han J.-W.* et al. Capacitive neural network with neuro-transistors // Nature Communications. 2018. Vol. 9. 3208.
33. *Zhang X.*, *Long K.* Improved learning experience memristor model and application as neural network synapse // IEEE Access. 2019. Vol. 7. Pp. 15262–15271.
34. *Minnekhanov A.A.*, *Emelyanov A.V.*, *Lapkin D.A.* et al. parylene based memristive devices with multilevel resistive switching for neuromorphic applications // Scientific Reports. 2019. Vol. 9. P. 10800.
35. *Liu Ju.*, *Harkin J.*, *Maguire L.P.* et al. Wade SPANNER: A self-repairing spiking neural network hardware architecture // IEEE Transactions on Neural Networks and Learning Systems. 2017. Vol. 29. No. 4. Pp. 1287–1300.
36. *Liu Ju.*, *Mcdaid L.J.*, *Harkin J.* et al. Exploring self-repair in a coupled spiking astrocyte neural network // IEEE Transactions on Neural Networks and Learning Systems. 2018. Vol. 30. No. 3. Pp. 865–875.

# ГЛАВА 1

# БИОМОРФНЫЙ НЕЙРОПРОЦЕССОР НА ОСНОВЕ ИНТЕГРАЦИИ ПЕРСПЕКТИВНЫХ НАНОРАЗМЕРНЫХ МЕМРИСТОРНЫХ ЭЛЕМЕНТОВ НАНОЭЛЕКТРОНИКИ С КЛАССИЧЕСКОЙ КМОП-НАНОТЕХНОЛОГИЕЙ — НОВОЕ НАПРАВЛЕНИЕ В ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ (IT)

## 1.1. МЕМРИСТОР В КАЧЕСТВЕ БЫСТРОДЕЙСТВУЮЩЕГО ПЕРЕКЛЮЧАТЕЛЯ В IT. МОДЕЛЬ ФОРМОВАННОГО МЕМРИСТОРА

Мемристор (*memory* + *resistor*) — это резистор, сопротивление которого изменяется при протекании тока через него, т. е., резистор, который «помнит» величину и длительность пропускаемого тока.

Мемристор представляет собой пассивный нелинейный элемент электрической цепи и является фундаментальным элементом электроники наравне с резистором, конденсатором и катушкой индуктивности (рис. 1.1) При протекании тока через мемристор в одном направлении его электрическое сопротивление увеличивается. Когда ток течет в обратном направлении — сопротивление уменьшается (рис. 1.2). В отсутствии электрического тока сопротивление мемристора сохраняется, и в момент подачи напряжения оно в точности равно сопротивлению до момента прекращения подачи питания. Следовательно, мемристор обладает энергонезависимой памятью.

Для создания электрических схем чипов памяти и аппаратных нейроморфных сетей, как правило, используются твердотельные мемристоры, изготовленные на тонких наноразмерных пленках и состоящие из оксидов переходных металлов. Физика процесса изменения проводимости таких мемристоров заключается в том, что при наложении электрического поля поперек пленки отрицательные ионы кислорода мигрируют к положительно заряженному электроду и обедняют объем вдали от электрода. Это приводит к росту проводимости мемристора, а при изменении полярности потенциала на электроде ионы кислорода из приэлектродной области движутся в обратную сторону и увеличивают сопротивление мемристора.

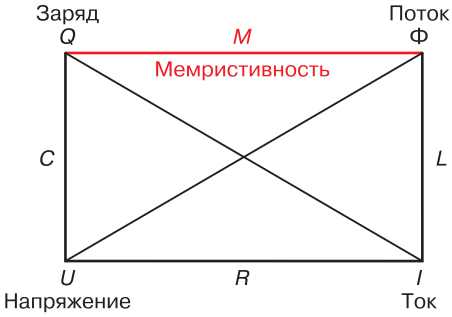



**Рис. 1.1.** Фундаментальные элементы электроники: конденсатор, резистор, индуктивность, мемристор

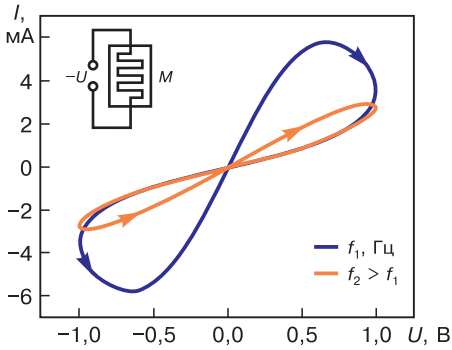



**Рис. 1.2.** Зависимость тока на мемристоре от приложенного напряжения

В аппаратных устройствах, на которых устанавливаются нейросети из простых нейронов и которые используются в информационных технологиях (IT), к твердотельным мемристорам предъявляются особые требования: стабильность и воспроизводимость их электрических параметров, а также малое время переключения из низкопроводящего в высокопроводящее состояние и наоборот. В современных мемристорах время переключения указанных состояний составляет единицы наносекунд.

Технически это достигается электрической формовкой мемристора при подаче на его электроды напряжения, значительно превышающего пороговое, необходимое для переключения состояний. В результате, например, диэлектрическая пленка из диоксида титана $T_iO_2$ из-за выноса ионов кислорода обедняется практически во всем своем объеме за исключением тонкого приэлектродного слоя толщиной несколько нанометров. Обедненный кислородом слой оксида титана $T_iO_{2-x}$ обладает сопротивлением порядка 215 Ом и по существу является проводником [1]. Между двумя разнородными по элементному составу электродами в диэлектрике при подаче напряжения имеет место туннелирование электронов через потенциальный

барьер [2]. Вольт-амперная характеристика в таком мемристоре зависит от полярности напряжения, подаваемого на электроды.

В работах [1; 3] развита модель туннелирования электронов через потенциальный барьер, в которой учитывается движение границы между диэлектриком и проводящим каналом (рис. 1.3).

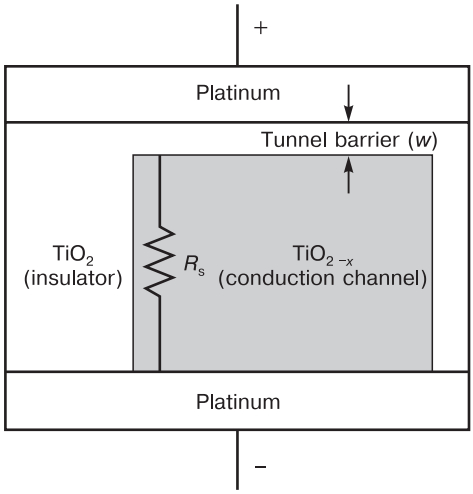



**Рис. 1.3.** Мемристор после процесса электроформования. Туннельный барьер между платиновыми электродами и проводящим каналом $T_iO_{2-x}$

Согласно этой модели, выражение для тока, протекающего через мемристор, имеет вид

$$i = \left(\frac{j_0 A}{\Delta w^2}\right)\left\{\varphi_1 \exp\left(-B\sqrt{\varphi_1}\right) - \left(\varphi_1 + e\left|v_g\right|\right)\exp\left(-B\sqrt{\varphi_1 + e\left|v_g\right|}\right)\right\};$$

$$j_0 = \frac{e}{2\pi h}; \quad w_1 = \frac{1{,}2\lambda w}{0}; \qquad w = w_2 - w_1;$$

$$\varphi_1 = \varphi_0 - e\left|v_g\right|\frac{w_1 + w_2}{w} - \frac{1{,}15\lambda w}{\Delta w}\ln\left[\frac{w_2\left(w - w_1\right)}{w_1\left(w - w_2\right)}\right];$$

$$B = \frac{4\pi\Delta w\sqrt{2m}}{h}; \quad \lambda = \frac{e^2\ln(2)}{8\pi k\varepsilon_0 w};$$

$$w_2 = w_1 + w\left(1 - \frac{9{,}2\lambda}{3\varphi_0 + 4\lambda - 2e\left|v_g\right|}\right),$$

где $w$ — ширина туннельного барьера; $A$ — площадь канала мемристора; $e$ — заряд электрона; $v_g$ — напряжение через туннельный барьер; $m$ — масса

электрона; $h$ —постоянная Планка; $k$ и $\varepsilon_0$ — диэлектрическая константа и проницаемость соответственно; $\varphi_0$ высота потенциального барьера в электронвольтах.

Изменение во времени ширины барьера $w$ (скорость движения границы между диэлектриком и проводящим каналом) описывается следующими выражениями:

- в случае переключения мемристора в низкопроводящее состояние ($i > 0$)

$$\frac{dw}{dt} = f_{off} \sinh\left(\frac{|i|}{i_{off}}\right) \exp\left[-\exp\left(\frac{w - a_{off}}{w_c} - \frac{|i|}{b}\right) - \frac{w}{w_c}\right];$$

- с подобранными параметрами $f_{off}$ =3,5 ± 1 мкс, $i_{off}$ = 115 ± 4 мкА, $a_{off}$ =1,2 нм, $b$ = 500 ± 70 мкА и $w_c$ = 107 ± 4 пкм; в случае переключения мемристора в высокопроводящее состояние ($i < 0$)

$$\frac{dw}{dt} = -f_{on} \sinh\left(\frac{|i|}{i_{on}}\right) \exp\left[-\exp\left(\frac{a_{on} - w}{w_c} - \frac{|i|}{b}\right) - \frac{w}{w_c}\right];$$

- с параметрами $f_{on}$ = 40 ± 10 мкс, $i_{on}$ =8,9 ± 0,3 мкА, $a_{on}$ = 1,8 ± 0,01 нм, $b$ = 500 ± 90 мкА и $w_c$ = 107 ± 3 пкм.

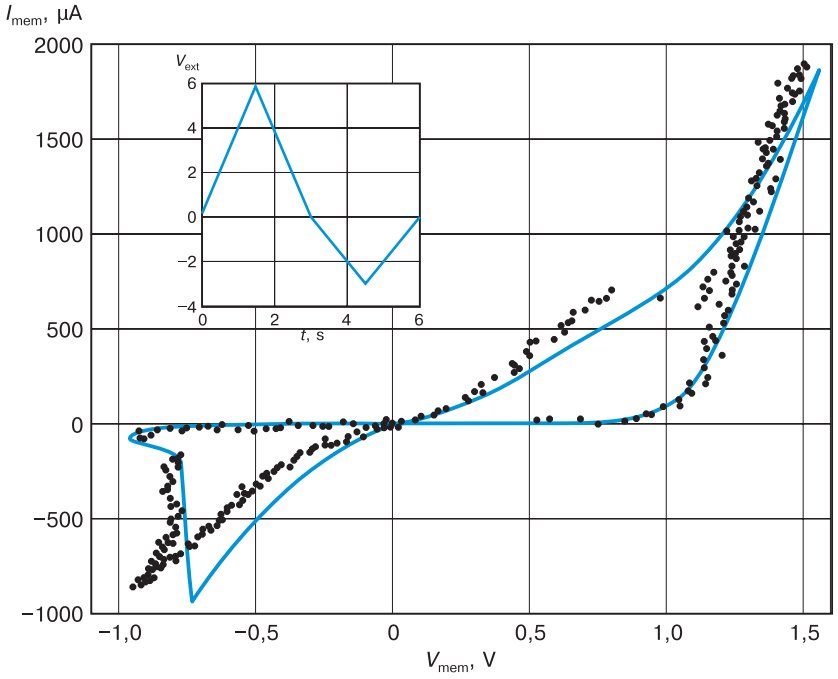



**Рис. 1.4.** Моделируемая вольт-амперная кривая для мемристора (*сплошная линия*) и соответствующие экспериментальные данные (*черные точки*). Форма импульса напряжения показана на врезке

На рис. 1.4 приведены вольт-амперная кривая меристора, которую дает SPICE-модель туннелирования электронов через потенциальный барьер шириной $w = 1{,}2$ нм, и вольт-амперная характеристика, полученная экспериментально при подаче напряжения +6/−3 В треугольной формы с периодом 6 с. Наблюдается хорошее качественное и количественное совпадение.

Моделируемая вольт-амперная кривая отличается от экспериментальных данных примерно на 20 %, поскольку большая нелинейность уравнений определяет высокую чувствительность к ошибкам в параметрах модели.

## 1.2. МЕМРИСТОР В КАЧЕСТВЕ СИНАПТИЧЕСКОЙ СВЯЗИ НЕЙРОНА С ДРУГИМИ НЕЙРОНАМИ

Если сопротивление синапса, связывающего два нейрона в биологической нейросети, возрастает при прохождении электрического импульса, то происходит процесс забывания, а если сопротивление уменьшается, происходит процесс запоминания информации в мозге. Твердотельный мемристор является аналогом биологического синапса, связывающего нейроны мозга [4; 5]. Мемристор действует как сопротивление, значение которого изменяется в зависимости от проходящего через него тока. В результате снижения сопротивления мемристора связь между логическими элементами, в формировании которой участвует этот мемристор, становится более эффективной — происходит обучение. При повышении сопротивления мемристора происходит разобучение (забывание). По своему действию мемристор подобен синапсу — соединению между нейронами мозга (рис. 1.5).

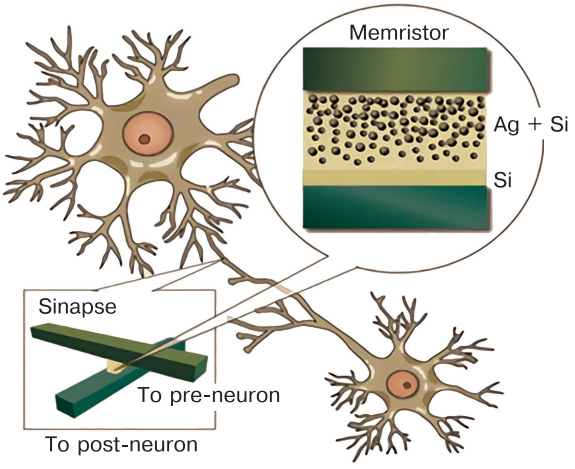



**Рис. 1.5.** Мемристор в качестве синаптической связи между нейронами

В работе [6] при исследовании вольт-амперных характеристик мемристора на основе смешанного оксида металлов $TiN/Ti_{0,92}Al_{0,08}O_{1,96}/TiN$, полученного магнетронным распылением двух катодов — мишеней Ti и Al в реактивной магнетронной среде, были обнаружены следующие эффекты.

### Релаксация состояния мемристора

При снятии напряжения с мемристорной структуры в результаты действия кулоновских сил ионы кислорода вытесняются из депо TiN, и их концентрация стремится к равновесной в толщине активной пленки. Время релаксации составляет от 0,8 секунды до десятков минут в зависимости от толщины слоя пленки TiN. Кроме того, время релаксации может быть дополнительно увеличено на несколько порядков введением между пленками $Ti_{0,92}Al_{0,08}O_{1,96}$ и TiN мембранного слоя TiO, затрудняющего диффузию вакансий [7].

Описанный эффект позволяет говорить о существовании кратковременной и долгосрочной памяти в полученном устройстве. О возможности существования эффекта кратковременной памяти в мемристорах сообщал Ting Chang в работах [8; 9], где также дано и обосновано сравнение принципов работы синапса и типового мемристора.

Показано, что при быстром повторном нагружении можно обратить процесс диффузии до достижения ионами кислорода равновесного состояния в тонкой пленке. На рис. 1.6, *1* — ВАХ первоначального открытия мемристора; на рис. 1.6, *2* и 1.6, *3* — ВАХ при подаче повторного зондирующего импульса. В случае *2* задержка больше, чем в *3*. На рис. 1.6, *4* — подача зондирующего импульса с минимальной задержкой. Описанный эффект может быть интерпретирован как обучаемость мемристора.

Обнаружен эффект сохранения достигнутого состояния при подаче импульсов напряжения с амплитудой меньше напряжения переключения, что соответствует импульсации в нервных клетках высших животных. Такие импульсы позволяют производить считывание состояния отдельного мемристора и сохранять в нем любое неограниченное время требуемое состояние.

График нелинейной ВАХ открытого мемристора (см. рис. 1.6, *1*) может быть апроксимирован с высокой точностью функцией гиперболического тангенса $thx = a(e^{2bx} - 1)/(e^{2bx} + 1)$, где $a$, $b$ зависят от свойств конкретного мемристивного элемента. Известно [10], что гиперболический тангенс используется в качестве суммирующей функции в математической модели искусственного нейрона с выходным сигналом, противостоящим насыщению.

Объединение мемристоров в кроссбары и построение с их помощью запоминающих и логических матриц дает возможность создать аппаратное средство — нейропроцессор. С помощью нейропроцессора и нейросети, установленной на нем, можно имитировать работу кортикальной колонки человеческого мозга. На рис. 1.7 представлены нейроморфные колонки мозга и мемристорная микросхема, соответствующая одному слою колонки.

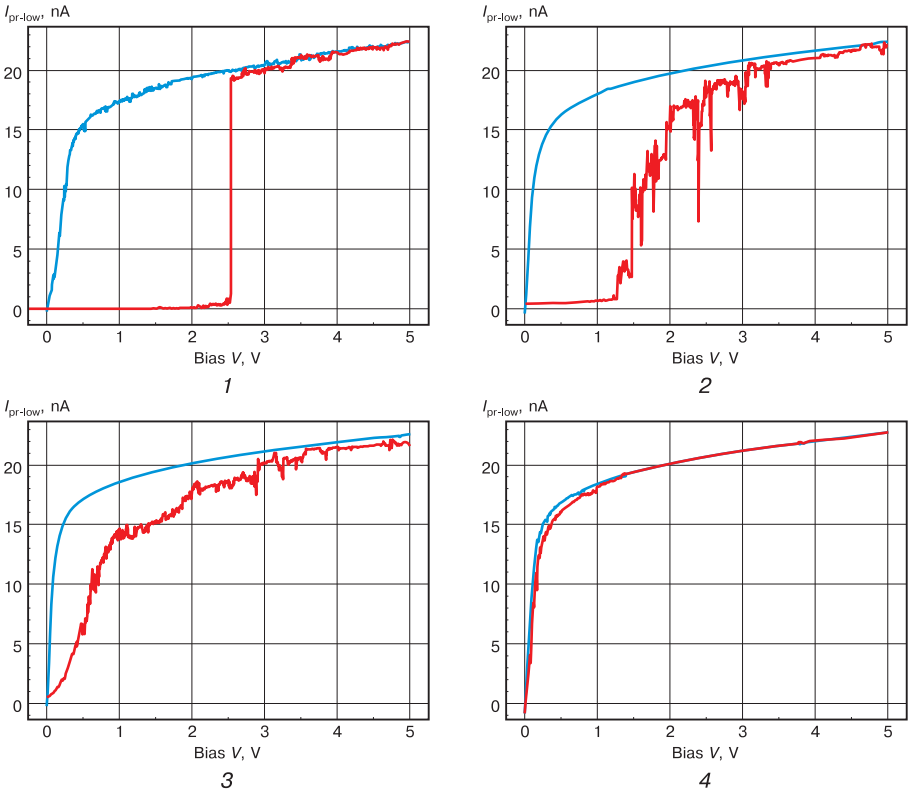



**Рис 1.6.** *1* — ВАХ первоначального открытия мемристора; *2*, *3*, *4* — ВАХ при подаче повторного зондирующего импульса с разной задержкой

В мемристоре между предельными высокопроводящим и низкопроводящим состояниями имеется множество промежуточных состояний с разной проводимостью. Эти состояния можно использовать в процессах ассоциативного обучения нейросети на основе мемристорных синапсов и одновременной обработки входных импульсов, заключающейся в их взвешивании и суммировании в аппаратном устройстве — нейропроцессоре. Время протекания этих процессов может значительно превосходить время переключения (порядка нескольких наносекунд [11]) формованного мемристора, используемого в качестве переключателя в информационных технологиях.

С этой точки зрения, формованные мемристоры с потенциальным барьером шириной порядка 1 нм непригодны. Необходимо использовать всю толщину пленки мемристора от 10 до 50 нм. При этом протекающий через неформованный мемристор электронный ток определяется медленным процессом диффузии отрицательных ионов кислорода. Так, расчет, проведенный в работе [12], показывает, что через мемристор толщиной $d = 10$ нм и отношением сопротивлений в высокоомном и низкоомном состояниях

$R_{off}/R_{on} = 160$ протекает электронный ток $i_0 = u/R_{on} = 10$ мА; при подвижности ионов $\mu_v = 10^{-10}$ см$^2$/с · В время переключения мемристора (миграции ионов через всю толщину пленки) составляет $t_0 = d^2/\mu_v u = 10$ мс.

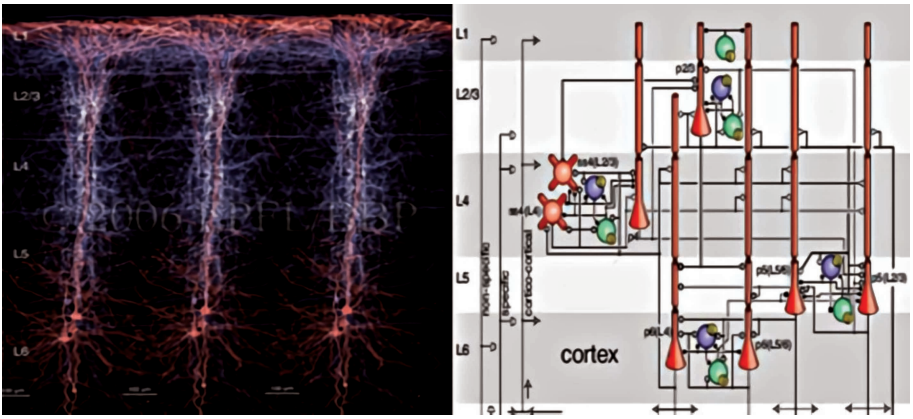



**Рис. 1.7.** Строение кортикальной колонки человеческого мозга. Мемристорная микросхема соответствует одному слою нейроморфной колонки

Это позволяет построить искусственные синапсы на основе мемристоров, которые будут работать с той же длительностью импульсов потенциала действия, что и в биологических нейронных сетях. Для работы биоморфного нейропроцессора будут применяться электрические импульсы, близкие по форме и длительности (порядка 2 мс) к биологическим импульсам, распространяющимся в кортикальной колонке мозга от нейрона к нейрону (рис. 1.8). Если мемристор полностью переключается за 10 мс, то для его полного переключения потребуется 5 импульсов по 2 мс.

На рис. 1.8 показана форма типичного потенциала действия на мембране нейрона. Потенциал остается вблизи базового уровня до тех пор, пока в какой-то момент времени он резко не поднимается вверх, а затем быстро падает. Приближенный график типичного потенциала действия показывает его различные фазы по мере того, как потенциал действия проходит точку на клеточной мембране. Мембранный потенциал покоя равен –70 мВ. Стимулирование начинается с 1 мс, когда мембранный потенциал достигает порогового значения –55 мВ, он быстро возрастает до пикового значения +40 мВ в момент времени 2 мс. Точно так же быстро потенциал затем падает до –90 мВ в момент времени 3 мс, и, наконец, потенциал покоя, равный –70 мВ, восстанавливается в момент времени 5 мс.

Использование неформованных мемристоров позволяет реализовать большое количество промежуточных резистивных состояний и, соответственно, большое количество возможных состояний синапсов нейронов, в отличие от быстро переключающихся формованных мемристоров, которые используются в цифровых запоминающих устройствах.

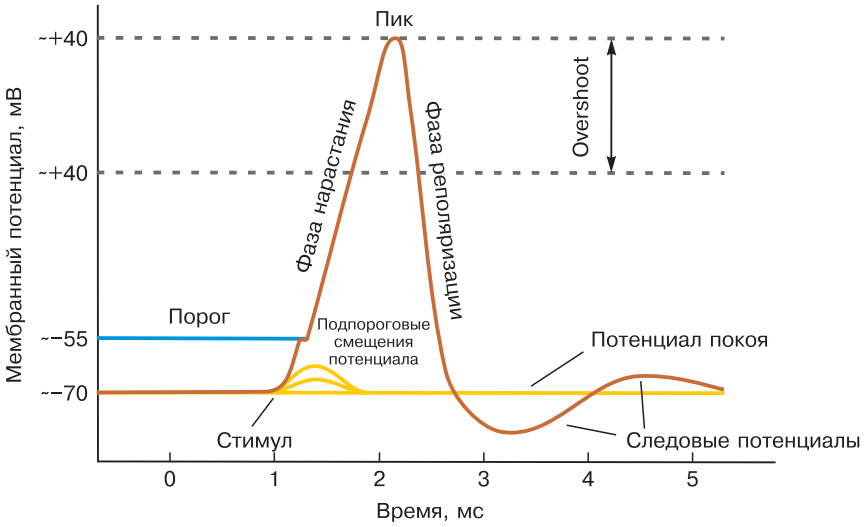



**Рис. 1.8.** Форма типичного потенциала на мембране нейрона

## 1.3. ФИЗИКО-МАТЕМАТИЧЕСКАЯ МОДЕЛЬ НЕФОРМОВАННОГО МЕМРИСТОРА

Для описания процесса резистивного переключения в мемристоре было предложено несколько физико-математических моделей. Среди них следует отметить модели электрополевой миграцией зарядов в объеме диэлектрика [12–14]. В [15] представлена модель резистивного переключения в мемристорном устройстве TiN/Ti/$HfO_2$/TiN. Ионные процессы были сымитированы с помощью кинетического метода Монте-Карло, уравнения Лапласа для поля, уравнения неразрывности для тока и уравнения теплопроводности. Рассмотрен один из возможных механизмов переноса тока электронами — туннелирование через ловушки (кислородные вакансии). Недостатком этой модели является то, что она привязана к особой структуре мемристора, в которой рождение и рекомбинация пары Френкеля (ион кислорода и кислородная вакансия) наблюдаются преимущественно на поверхности титанового электрода. Это происходит благодаря высокому сродству титана к кислороду. Энергия активации для образования пары Френкеля оказывается значительно меньше на границе титанового электрода, чем в объеме оксида. В мемристорных структурах с электродами, обладающими низким сродством к кислороду, следует рассматривать объемные процессы рождения и рекомбинации пар Френкеля. В [16] описано туннелирование в структуре МОМ захваченных электронов через вакансии кислорода с последующей релаксацией атомного окружения этих вакансий.

В работе [17] сформулирована наиболее полная нелинейная модель тепломассопереноса зарядов в структуре металл–оксид–металл при доминирующем транспортном механизме туннелирования электронов через кислородные вакансии, с помощью которой возможно построение модели, описывающей различные резистивные состояния и переключения мемристора из высокопроводящего в низкопроводящее состояние и наоборот. Термодинамическая модель учитывает процессы рождения и рекомбинации пар Френкеля в объеме оксида и включает нестационарное уравнение теплопроводности с нелинейным источником тепла, содержащим джоулев нагрев и рекомбинацию зарядов, уравнение Пуассона для электрического поля и уравнений Шокли–Рида–Холла с учетом сильного электрон-фононного взаимодействия в процессах ионизации ловушек и переноса заряда в объеме оксида металла. Предложенные в работах [16; 17] физические модели переноса зарядов в оксиде металла достаточно сложны и требуют ресурсоемких численных квантово-химических расчетов, за которыми скрывается физический смысл протекающих процессов.

В [18] представлена математическая модель резистивных состояний и динамического переключения мемристора на основе более простой физической модели массопереноса зарядов без учета процесса теплопереноса в структуре металл–оксид–металл при доминирующем транспортном механизме туннелирования электронов через кислородные вакансии, мигрирующие под действием неоднородного самосогласованного электрического поля. Для этого реализована модель мемристора в виде специализированной программы и проведено численное моделирование состояний мемристора, а также сравнение полученных в результате расчета вольт-амперных характеристик мемристора с экспериментальными данными.

### 1.3.1. Математическая модель переноса зарядов в мемристоре

Нестационарные одномерные уравнения массопереноса кислородных вакансий и отрицательных ионов кислорода включают члены, описывающие процессы генерации, рекомбинации и миграции в электрическом поле [17]:

$$\frac{\partial N}{\partial t} = \left(N_0 - N\right)\frac{W_{\text{ph}}}{h}\exp\left[\frac{-\left(W_{\text{ox}} - \sqrt{\frac{q^3 E}{\pi\varepsilon\varepsilon_0}}\right)}{kT}\right] - NN_{\text{ox}}v_{\text{ox}}\sigma_{\text{ox}} - \\ -s_0\frac{\partial}{\partial z}\left\{N\left(1-\frac{N}{N_0}\right)\frac{W_{\text{ph}}}{h}\exp\left[\frac{-\left(W_{\text{ox}} - \frac{q^2}{\pi\varepsilon\varepsilon_0 s_0}\right)}{kT}\right]\sinh\left(\frac{qEs_0}{2kT}\right)\right\}; \quad (1.1)$$

$$\frac{\partial N_{\text{ox}}}{\partial t} = \left(N_0 - N\right)\frac{W_{\text{ph}}}{h}\exp\left[\frac{-\left(W_{\text{ox}} - \sqrt{\frac{q^3 E}{\pi\varepsilon\varepsilon_0}}\right)}{kT}\right] - NN_{\text{ox}}v_{\text{ox}}\sigma_{\text{ox}} + \frac{\partial}{\partial z}\left(N_{\text{ox}}v_{\text{ox}}\right), \quad (1.2)$$

где $N$ — концентрация кислородных вакансий (ловушек); $N_{max}$ — максимально возможная концентрация вакансий; $N_{ox}$ — концентрация междоузельных ионов кислорода; $s_0 = N_{max}^{-1/3}$ — среднее расстояние между атомами; $W_{ph}$ — энергия фонона; $W_{ox}$ и $\sigma_{ox}$ — энергия образования и рекомбинационное сечение для междоузельной-вакансионной пары; $E$ — электрическое поле; $v_{ox}$ — скорость дрейфа междоузельных атомов; $T$ — температура диэлектрика; $z$ — координата поперек пленки оксида металла; $h$ — постоянная Планка; $k$ — постоянная Больцмана; $q$ — заряд вакансии; $\varepsilon$ — низкочастотная диэлектрическая проницаемость среды.

Первый член в правой части уравнения (1.1) описывает скорость генерации новой пары междоузельный ион — вакансия в сильном электрическом поле по механизму Френкеля [19]. Второй член в (1.1) описывает рекомбинацию кислородных вакансий с междоузельными ионами кислорода. Третий член связан с дрейфом вакансий кислорода, который происходит в виде прыжков соседних ионов кислорода на позиции вакансий по механизму Хилла [20]. При этом первоначальные вакансии рекомбинируют, а на месте ионов образуются новые вакансии. Уравнение (1.2) описывает аналогичные процессы для междоузельных ионов кислорода. Первый и второй члены в правой части (1.2) аналогичны соответствующим членам в уравнении (1.1). Третий член описывает дрейф междоузельных ионов кислорода в электрическом поле.

Численное моделирование в [21] процесса генерации кислородных вакансий в структуре $Pt/HfO_2/TiN$ при толщине оксидного слоя $d = 3$ нм и напряжении на электродах $u = 1$ В показало, что концентрация вакансий выходит на постоянное значение через время $t = 0{,}1$ мс. Время медленного процесса миграции отрицательных ионов кислорода по толщине оксидного слоя $d = 10$ нм в структуре $Pt/TiO_x/Pt$ с подвижностью $\mu_v \approx 10^{-10}$ см$^2$/с · В [12] составляет порядка $t_0 \approx d^2/\mu_v u = 10$ мс. Время переключения структуры из низкопроводящего в высокопроводящее состояние является нелинейной функцией амплитуды приложенного напряжения. Так, время переключения структуры $Pt/HfO_2/Ti$ при напряжении до 3В составило $5 \cdot 10^{-2}$ мс, а в структуре $Pt/TiO_x/Pt$ при напряжении до 1 В, соответственно, $10^{-1}$ мс [22]. Следовательно, уравнения (1.1) и (1.2), определяющие концентрации ионов и вакансий, можно рассматривать в стационарном режиме. В результате сложения этих уравнений для концентрации вакансий получим следующее нелинейное дифференциальное уравнение первого порядка с переменным коэффициентом $C(z)$:

$$\frac{\partial N(z)}{\partial z} - \sigma_{ox} N^2(z) - C(z)\left[N_{max} - N(z)\right] = 0;$$

$$C(z) = \exp\frac{\left[\left(\sqrt{\frac{q^3 E(z)}{\pi\varepsilon\varepsilon_0}} - \frac{q^2}{\pi\varepsilon\varepsilon_0 s_0}\right)\right]}{kT} \sinh\left(\frac{qE(z)s_0}{2kT}\right). \qquad (1.3)$$

Поскольку в диэлектрике с низкой электропроводностью плотность тока свободных электронов мала и доминирующим транспортным механизмом является туннелирование электронов через мигрирующие в электрическом поле вакансии [15], представим плотность тока через оксидный слой мемристора в виде [17; 23]:

$$J = \frac{e}{s^2}\frac{n_t}{N}\left(1-\frac{n_t}{N}\right)P_{\text{tun}}, \tag{1.4}$$

где $n_t$ — концентрация захваченных в ловушки (вакансии) электронов; $s = N^{-1/3}$ — среднее расстояние между ловушками; $P_{\text{tun}}(u)$ — частота туннелирования электронов между ловушками.

Частота туннелирования между фонон-связанными ловушками имеет экспоненциальную зависимость от электрического поля и температуры:

$$P_{\text{tun}} = \frac{2\sqrt{\pi}\hbar W_t}{m^* s^2 Q_0 \sqrt{kT}}\exp\left(-\frac{W_{\text{opt}}-W_t}{2kT}\right)\exp\left(-\frac{2s\sqrt{2m^*W_t}}{\hbar}\right)\sinh\left(\frac{esE}{2kT}\right), \tag{1.5}$$

где $Q_0 = \sqrt{2\left(W_{\text{opt}}-W_t\right)}$, $W_t$ и $W_{\text{opt}}$ — термическая и оптическая энергия ионизации ловушки; $m^* = 0{,}25m_{\text{e}}$ — эффективная масса носителя заряда; $m_{\text{e}}$ — масса электрона; $\hbar$ — приведенная постоянная Планка.

Концентрация захваченных в ловушки (вакансии) электронов, входящая в формулу (1.4), в стационарных условиях соответствует статистике Ферми на контакте металл—диэлектрик [17]

$$n_t(z=0) = \frac{N}{1+\exp\left(\dfrac{A_M-\chi_D-W_t-eu}{kT}\right)}, \tag{1.6}$$

где $A_M$ — работа выхода электрона из металла в вакуум; $\chi_D$ — электронное сродство в диэлектрике; $eu$ — приобретенная энергия электрона в электрическом поле.

Таким образом, в формуле (1.1) отражено уменьшение потенциального барьера на контакте металл–диэлектрик при приложении внешней разности потенциалов на электроды [24].

### 1.3.2. Аналитическая и численная модели переключения мемристора

Из-за малой электропроводности доминирующим транспортным механизмом электронов в слое оксида металла является туннелирование через ловушки. В этом случае внешнее электрическое поле $E$ незначительно искажается свободными и захваченными носителями заряда, и его

стационарное распределение в диэлектрическом слое определяется с помощью уравнения Лапласа [15]:

$$\frac{\partial^2 \varphi}{\partial z^2} = 0; \quad E = -\frac{\partial \varphi}{\partial z} = -\frac{u}{d}; \tag{1.7}$$

$$\varphi\big|_{z=0} = 0; \quad \varphi\big|_{z=d} = u, \tag{1.8}$$

где $d$ — толщина слоя оксида металла; $u$ — разность потенциалов (напряжение) на электродах мемристора.

В этом случае уравнение (1.3) с постоянным коэффициентом $C$ для концентрации вакансий допускает аналитическое решение

$$N(z,\, u) = \frac{C(u)}{2\sigma_{ox}(u)} + \frac{D(u)}{2\sigma_{ox}(u)} \times \\ \times \mathrm{tg}\left[\frac{D(u)}{2}(z-d) + \mathrm{arc\,tg}\left(\frac{2\sigma_{ox}(u) N_{z=d} - C(u)}{D(u)}\right)\right], \tag{1.9}$$

где $D = \sqrt{4\sigma_{ox} N_0 C - C^2}$ .

Процессом рекомбинации кислородных вакансий по сравнению с их дрейфом под действием электрического поля в уравнении (1.3) можно пренебречь, если $N^2 \sigma_{ox}/N_{max} \ll C$, где $\sigma_{ox} = e/4\varepsilon\varepsilon_0 E$ [17]. Это условие хорошо выполняется при невысоких температурах и большом электрическом поле. Решение уравнения (1.3) в таком случае имеет вид:

$$N(z,\, u) = N_0 - \left(N_0 - N_{z=d}\right)\exp\left[-C(u)(d-z)\right]. \tag{1.10}$$

Концентрация захваченных электронов с увеличением напряжения на электродах приближается к концентрации вакансий, что приводит к ограничению роста плотности тока (1.4) при напряжении, равном напряжению переключения мемристора из низкопроводящего в высокопроводящее состояние. Вблизи напряжения переключения мемристора справедливо уравнение Лапласа для электрического поля (1.7). При меньших напряжениях вместо этого уравнения необходимо решать уравнение Пуассона для электрического поля, в правой части которого присутствует разность плотности зарядов вакансий и электронов $e(N - n_t)$.

Поскольку определение концентрации захваченных в ловушки электронов является сложной задачей [17], найдем величину электрического поля из условия равенства диффузионного потока вакансий и их потока, обусловленного электрическим полем, т. е. при равновесном распределении вакансий [25]

$$E(z) = -\frac{\partial \varphi(z)}{\partial z} = \frac{T}{qN(z)} \frac{\mathrm{d}N(z)}{\mathrm{d}z}, \tag{1.11}$$

где $\varphi(0) = 0$; $\varphi(d) = u$.

Это условие следует из стационарного уравнения дрейфа–диффузии вакансий при постоянных коэффициенте диффузии и температуре [26]. В такой постановке математической модели мемристора ставится задача о численном решении дифференциального уравнения первого порядка с переменным коэффициентом $C(z)$ (1.3).

Искомый ток, который протекает через оксидный слой и измеряется на электроде с нулевым потенциалом (в месте эмиссии электронов), можно рассчитать аналитически с помощью выражений (1.4)–(1.9), определяя концентрации вакансий и инжектированных электронов у этого электрода с известной площадью. При численном расчете тока через мемристор для определения концентрации кислородных вакансий на границе электрода необходимо решать нелинейную систему уравнений (1.3) и (1.11).

## 1.4. ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ ПРОЦЕССОВ В МЕМРИСТОРЕ

### 1.4.1. Реализация модели мемристора в виде программы

Программа для расчета резистивных состояний и переключения мемристора составлена на языке программирования Python версии 3.7 с применением библиотеки numpy. Значения всех переменных были представлены типом numpy.double, который соответствует типу double в языке C. При расчетах использовалась расчетная сетка с равномерным расположением узлов.

Процедура расчета профиля вакансий кислорода при определенном напряжении на электродах состоит из трех этапов. В качестве начального приближения распределение вакансий вычисляется при постоянном электрическом поле по толщине оксидного слоя мемристора. Затем рассчитывается электрическое поле, соответствующее найденному скалярному полю вакансий. Далее, на основе получившегося электрического поля вычисляется итоговый профиль концентраций вакансий по толщине пленки. Расчет электрического поля, соответствующего определенному профилю вакансий кислорода, осуществляется с использованием формулы (1.11). При этом для численного вычисления производной концентрации вакансий по толщине слоя применяется двусторонняя конечная разность

$$\left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)_i = \frac{N_{i+1} - N_{i-1}}{z_{i+1} - z_{i-1}}, \quad i = 1\ldots(M-1). \tag{1.12}$$

На границах слоя диэлектрика использованы соответствующие односторонние конечные разности

$$\left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)_0 = \frac{N_1 - N_0}{z_1 - z_0}, \quad \left(\frac{\mathrm{d}N}{\mathrm{d}z}\right)_M = \frac{N_M - N_{M-1}}{z_M - z_{M-1}}. \tag{1.13}$$

Скалярное поле концентраций вакансий рассчитывается путем численного решения задачи Коши для уравнения (1.3) методом Эйлера. Подставляя одностороннюю конечную разность в уравнение (1.3), получим соответствующую итерационную формулу

$$N_{i+1} = N_i + \left(z_{i+1} - z_i\right)\left(\sigma_{ox} N_i^2 + C\left(z_i\right)\left[N_{max} - N_i\right]\right). \qquad (1.14)$$

Концентрация вакансий $N_0$ для начальной итерации соответствует концентрации вакансий в отсутствие электрического поля. Для уменьшения погрешности при численном расчете была выполнена процедура обезразмеривания путем следующих замен

$$N = nN_{max}; \quad z = xd.$$

Плотность тока рассчитывается по формулам аналитической модели с использованием скорректированного значения электрического поля вблизи катода. Построение вольт-амперной характеристики выполняется путем последовательного расчета профиля вакансий и соответствующей плотности тока при разных напряжениях.

### 1.4.2. Моделирование резистивных состояний и переключения мемристора

Для аналитического и численного моделирования расчетные параметры ловушек в пленке оксида гафния $HfO_2$ взяты из работы [23]: $W_t$ = 1,25 эВ, $W_{opt}$ = 2,5 эВ. Параметры, связанные с геометрией мемристора, позаимствованы из [27], в которой исследована структура $Pt/HfO_2/TiN$ с толщиной оксидного слоя 6 нм и площадью электродов $2 \times 2$ мкм$^2$.

На рис. 1.9 показано распределение электрического поля по толщине оксидного слоя мемристора, используемое при аналитическом и численном моделировании.

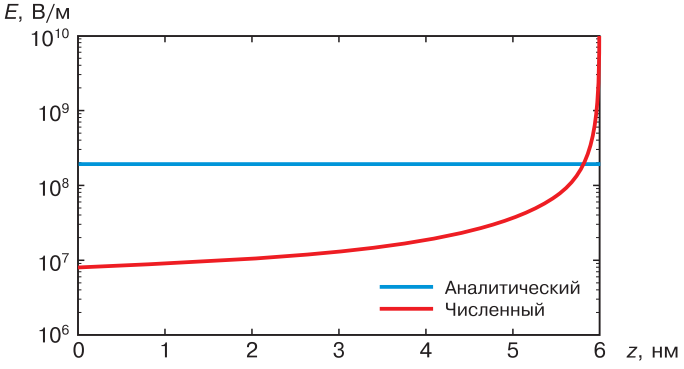



**Рис. 1.9.** Распределение электрического поля по толщине оксидного слоя мемристора согласно формулам (1.7) и (1.11) для аналитического и численного моделирования соответственно

На рис. 1.10 представлены распределения концентрации кислородных вакансий по толщине пленки оксида гафния $HfO_2$ в мемристоре при разной температуре.

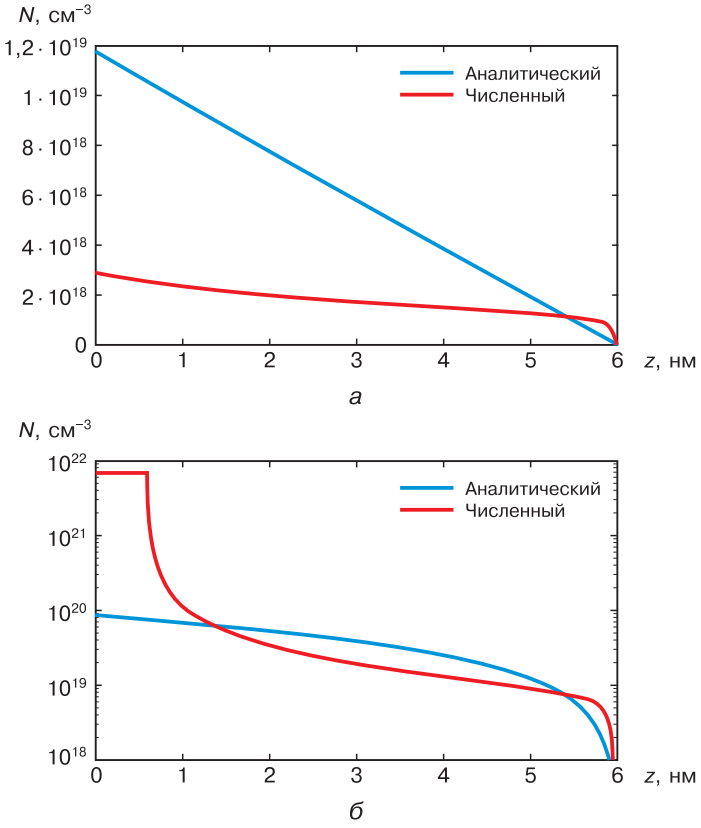



**Рис. 1.10.** Распределение концентрации кислородных вакансий по толщине пленки оксида гафния $HfO_2$:
*а* — 200 °C; *б* — 313 °C

Из рис. 1.10 следует, что большей температуре вакансий соответствует более высокая их концентрация. С ростом концентрации вакансий увеличивается электронный ток через мемристор, что в свою очередь увеличивает тепловыделение в результате омического нагрева в локальной области оксида, в которой протекает ток. Результаты измерений в [27] свидетельствуют о том, что в мемристоре выделяется мощность порядка 100 мкВт и в области протекания тока температура достигает значений более 630 °C. Согласно рис. 1.10, *б*, численная модель, учитывающая самосогласованное электрическое поле, более точно описывает процесс увеличения и выхода на полку концентрации вакансий с ростом температуры вблизи электрода

с отрицательным потенциалом. Подобный равновесный профиль концентрации вакансий получен в работе [28].

На рис. 1.11 можно увидеть свидетельство высокой зависимости максимального значения концентрации вакансий на границе с электродом от приложенного напряжения.

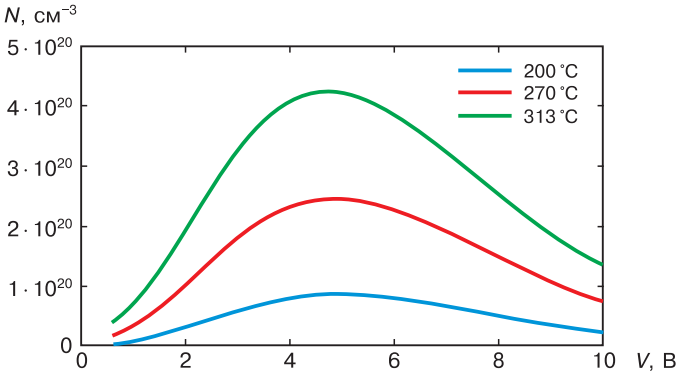



**Рис. 1.11.** Зависимость концентрации вакансий от приложенного напряжения при разных температурах вблизи электрода с низким потенциалом

На рис. 1.12 построены зависимости концентрации вакансий и инжектируемых в оксид электронов, захватываемых ловушками, от приложенного напряжения на электродах мемристора.

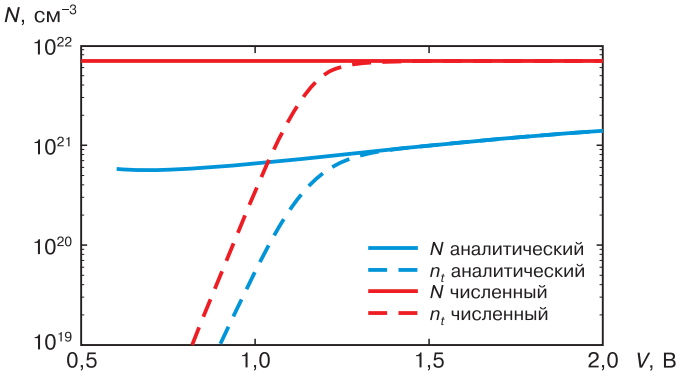



**Рис. 1.12.** Зависимости концентрации кислородных вакансий и инжектированных электронов от приложенного напряжения при температуре 313 °C.

Из этой зависимости следует, что с ростом напряжения происходит выравнивание концентраций захваченных электронов и вакансий кислорода, что должно в итоге привести к насыщению тока. Математически это следует

из формулы (1.4): экспоненциальный рост частоты туннелирования электронов через ловушки компенсируется уменьшением множителя $(1 - n_t/N)$.

На рис. 1.13 представлены расчетные и экспериментальные вольт-амперные характеристики (ВАХ): *а* — для структуры $Pt/HfO_2/TiN$ [27] и *б* — для $TiN/HfO_2/Ti/TiN$ [29], в которой толщина оксидного слоя 10 нм и площадь электродов $1 \times 1$ мкм$^2$. Разработанные аналитическая и численная модели мемристора позволяют определить температуру в мемристоре, при которой достигается совмещение площадей ВАХ и совпадение соответствующих параметров мемристора в низкопроводящем и высокопроводящем состояниях.

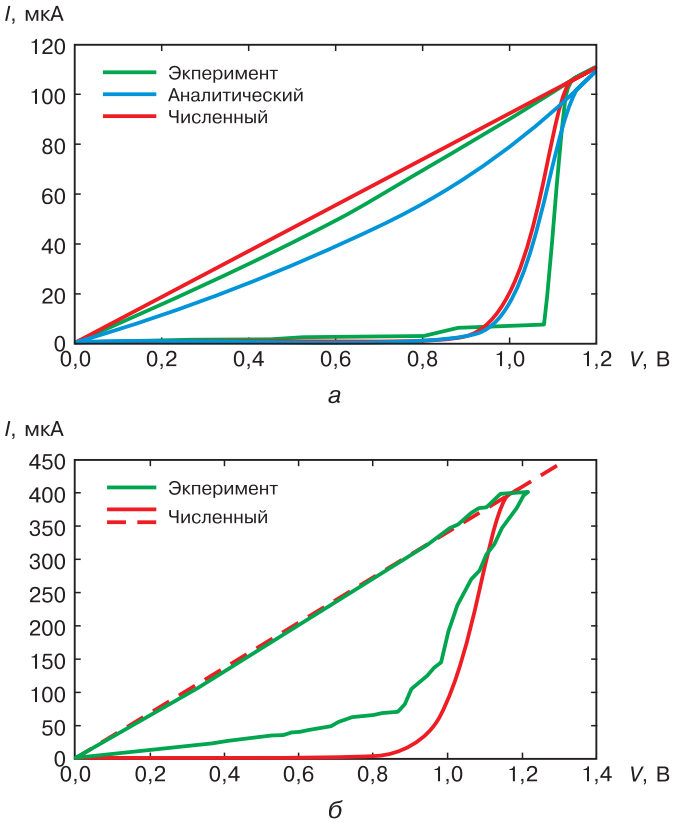



**Рис. 1.13.** Расчетные и экспериментальные ВАХ:
*а* — мемристора [18], аналитический расчет при 435 °С и численный — при 313 °С;
*б* — мемристора [19], численный расчет при 434 °С

Сравнение ВАХ на рис. 1.13, *а* показывает, что численное моделирование переключения мемристора из низкопроводящего в высокопроводящее состояние с учетом самосогласованного электрического поля дает более точное совпадение расчетной и экспериментальной ВАХ по сравнению с аналитическим приближением.

Несовпадение расчетной и экспериментальной кривых при увеличении тока в мемристоре [29] на рис. 1.13, *б* связано с высокой температурой в области протекания тока с повышенной концентрацией электронов. При выделяемой мощности 400 мкВт средняя температура оксидной пленки из-за теплопереноса поперек и вдоль направления тока составила всего 140 °C.

Для построения более точной математической модели мемристора необходимо включать в нее расчет теплопереноса на основе трехмерного нелинейного дифференциального уравнения второго порядка. Отметим также, что представленная модель мемристора не учитывает структуру и элементный состав электродов мемристора, различные приэлектродные процессы рождения и рекомбинации ионов кислорода и вакансий. Влияние материала электродов в модели описывается только работой выхода электронов.

Если подавать на электроды напряжение меньше, чем критическое напряжение переключения в крайнее высокопроводящее состояние, то можно достигать разные дискретные состояния мемристора. На практике этого добиваются ограничением тока во внешней цепи, путем последовательного с мемристором включения токоограничивающего резистора. Кривые *1–4* на рис. 1.14 описывают дискретные состояния мемристора [27] с разной величиной проводимости.

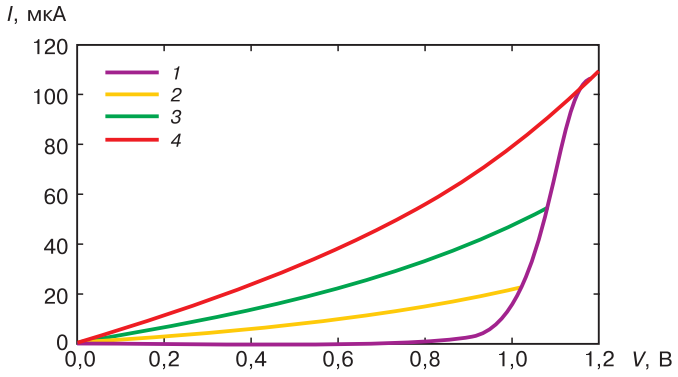



**Рис. 1.14.** Дискретные состояния мемристора:
*1* — закрытое; *2* — при 65 % от максимума вакансий;
*3* — при 84 % от максимума вакансий;
*4* — минимальное сопротивление (максимум вакансий)

Таким образом, создана математическая модель резистивных состояний и динамического переключения мемристора из низкопроводящего в высокопроводящее состояние на основе физической модели массопереноса зарядов без учета процесса теплопереноса в структуре металл–оксид–металл при доминирующем транспортном механизме туннелирования электронов через кислородные вакансии, мигрирующие под действием неоднородного самосогласованного электрического поля. В приближении постоянного электрического поля реализована аналитическая модель мемристора

с активным слоем из оксида переходного металла. Представлена численная модель мемристора в виде специализированной программы на основе метода конечных разностей, позволяющей учесть неоднородность самосогласованного электрического поля для определения его влияния на электрические характеристики мемристора.

Найдены распределения концентраций вакансий по толщине оксида и в зависимости от приложенного напряжения на электродах. Показано, что насыщение тока происходит при достижении напряжения переключения мемристора, когда сравниваются концентрации электронов и ловушек вблизи электрода с отрицательным потенциалом. При этом наблюдается стабилизация неустойчивости — неограниченного роста электронного тока, обусловленной экспоненциальным ростом частоты туннелирования электронов между фонон-связанными ловушками при увеличении напряжения на электродах.

Построенная с помощью разработанной математической модели переключения мемристора вольт-амперная характеристика согласуется с экспериментальными данными. Численную модель можно применять при исследовании и разработке мемристоров с заданными электрическими характеристиками.

Модель динамического переключения мемристора описывает дискретные состояния мемристора с разной величиной проводимости, что позволяет применять ее для моделирования работы устройств искусственного интеллекта, использующих мемристоры в качестве синапсов нейронов. Модель, построенная на основе простого аналитического выражения для электронного тока, может эффективно быть задействована при оптимизации процессов обработки сигналов в сверхбольших запоминающей и логической матрицах нейропроцессора, созданных на мемристорных кроссбарах.

## 1.5. ИНТЕГРАЦИЯ МЕМРИСТОРНЫХ УСТРОЙСТВ С КМОП-ЛОГИКОЙ

Аппаратная реализация нейросетей очень важна для их дальнейшего использования. Однако современная коммерческая электроника (основанная на КМОП — комплементарных металл–оксид–полупроводник структурах) не позволяет быстро и эффективно производить нейросетевые вычисления. Таким образом, интеграция нейроморфных устройств с коммерческой электроникой является важным направлением исследований.

Нейроморфные КМОП-устройства уже разрабатываются, например, процессор IBM TrueNorth [30] благодаря многоядерной архитектуре обеспечивает достаточную производительность для моделирования кортикальной колонки мозга. Каждое из 4096 ядер моделирует 256 нейронов с 256 синапсами у каждого нейрона, что достаточно для функционального

моделирования одного среза кортикоморфной колонки. Тем не менее подобная реализация использует 5,4 млрд транзисторов и энергозависимую память. Таким образом, масштабирование такой системы приведет к существенному увеличению энергопотребления и замедлению вычислений. Использование мемристоров в качестве синаптических связей в подобных устройствах позволит упростить их архитектуру, что увеличит быстродействие и снизит потребление энергии.

Интегрирование мемристивных устройств с КМОП-логикой уже применяется при разработке аппаратных нейроморфных сетей [31] и создании чипов памяти [32; 33]. Возможность использования мемристоров в качестве синапсов искусственных нейросетей подтверждена экспериментально [34].

Электронное устройство, представленное в [35], включает в себя микропроцессорную реализацию тела нейрона, интегрированную с кроссбаром мемристоров, проводники которого выступают в роли аксонов и дендритов, а сами мемристоры имитируют биологические синапсы. Такая аналоговая структура способна к ассоциативному запоминанию и объединяет память и процессинг в одном электронном устройстве.

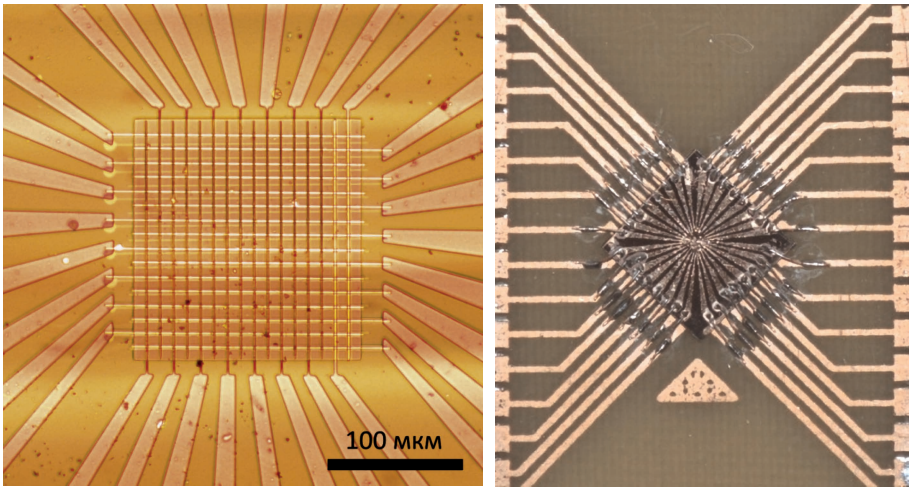
Мемристорная микросхема выполнена по технологии crossbar на основе слоя смешанного оксида металлов. В магнетронном модуле нанотехнологического комплекса «НаноФаб-100» НТ МДТ методом распыления двух катодов из алюминия и титана в среде кислорода получена композитная тонкопленочная структура $TiN/Ti_{0,92}Al_{0,08}O_{1,96}/TiN$ [6]. С помощью маски, выполненной методом электронно-лучевой литографии, сверху и снизу пленки смешанного оксида нанесены 16 проводящих дорожек шириной 400 нм. Толщина слоя смешанного оксида и электродного подслоя нитрида титана составила 30 нм. Как показано в [36], слой из нитрида титана между алюминиевым электродом и активной пленкой является эффективным резервуаром кислорода, а также источником дополнительных вакансий кислорода за счет образования дополнительного граничного TiNO подслоя.

КМОП-часть устройства построена на двух микроконтроллерах семейства PIC18, связанных шиной $I^2C$, с возможностью подключения к компьютеру посредством USB. Каждый контроллер соединен со своей стороной кроссбара ($2 \times 64$ мемристоров) независимо. Остальные мемристоры задействуются при одновременной работе двух микроконтроллеров. На рис. 1.15 представлены кроссбар и мемристорная микросхема, а на рис. 1.16 — лабораторная системная плата электронного устройства.

Снятие кривых ВАХ выполнено посредством атомно-силового микроскопа NT-MDT «NTegra». Встроенные источник постоянного напряжения и осциллограф позволяют производить измерения в диапазоне −10–10 В с точностью 10 пА.

На рис. 1.17 представлена вольт-амперная характеристика мемристора в кроссбаре $TiN/Ti_{0,92}Al_{0,08}O_{1,96}/TiN$

Стабильность электрических характеристик полученной мемристорной структуры выше, чем созданной с помощью атомно-слоевого осаждения структуры $Ti/Ti_{0,85}Al_{0,15}O_{1,93.}/Pd$, в которой толщина активного слоя такая же, 20 нм [37].




*а* *б*

**Рис. 1.15.** Кроссбар (*а*) и мемристорная микросхема (*б*)

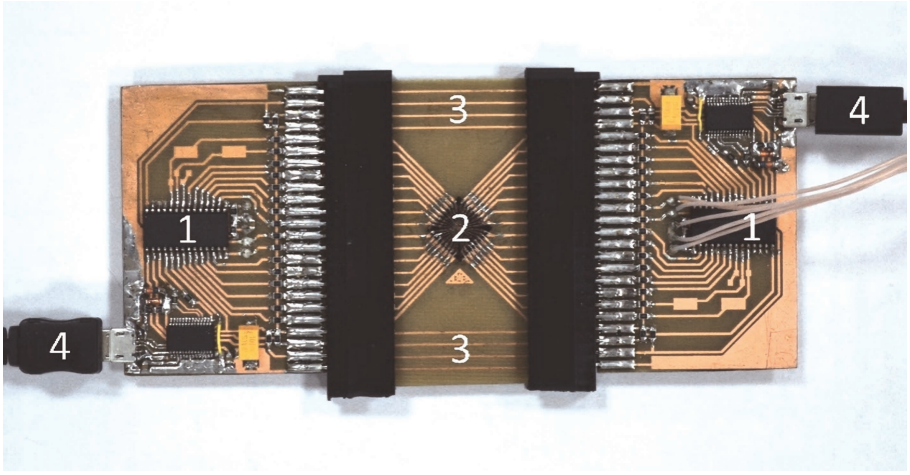



**Рис. 1.16.** Лабораторная системная плата:
1 — PIC18-контроллер; 2 — мемристорная микросхема;
3 — $I^2C$-интерфейс; 4 — USB-интерфейс

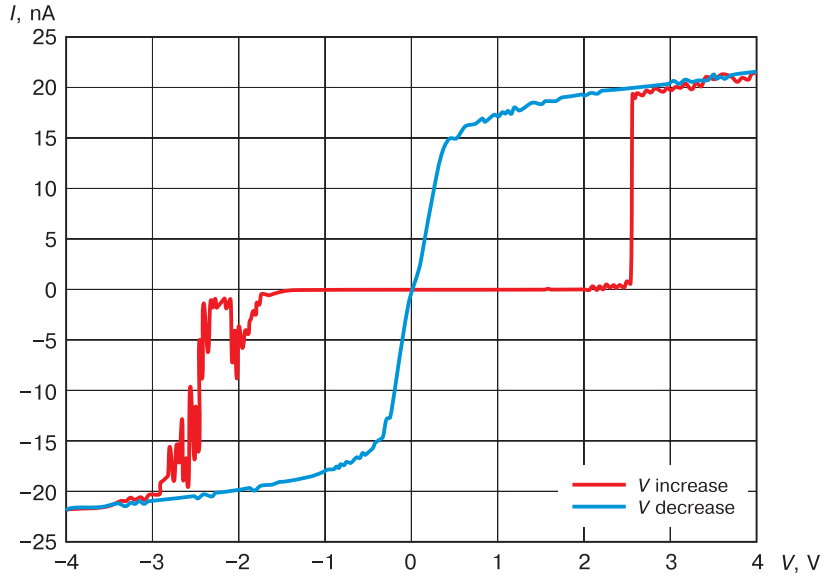



**Рис. 1.17.** Вольт-амперная характеристика мемристора на основе смешанного оксида металлов

Среднеквадратичное отклонение сопротивления в высокопроводящем состоянии первой структуры составило 3,3 % при напряжении переключения 2,69 ± 0,18 В (рис. 1.18), в то время как аналогичное отклонение во второй структуре составило 5,5 % при напряжении переключения 2,1 ± 0,2 В. Как следует из рис. 1.18, отношение сопротивлений в низкопроводящем и высокопроводящем состояниях структуры, полученной магнетронным распылением, составляет более 700. Это отношение сопротивлений должно быть максимально большим, поскольку для нейроморфных вычислений важно существование множества промежуточных синаптических состояний [38].

На микропроцессорах для упрощения ввода данных была установлена простая нейросеть — однослойный персептрон. Созданное электронное устройство является важным шагом на пути к нейроморфному процессору с различными функциями — компьютерного зрения, слуха, чтения текстов и т.д. Установленную на электронном устройстве нейросеть можно использовать как интерфейс по сбору, обработке и передаче данных. Усложнив программу дополнительными слоями нейронов (тремя слоями и более), можно получать более высокие ассоциации сенсорных данных — сигналов с матрицы видеокамеры или со звукового сигнального процессора, выполняющего преобразование Фурье.

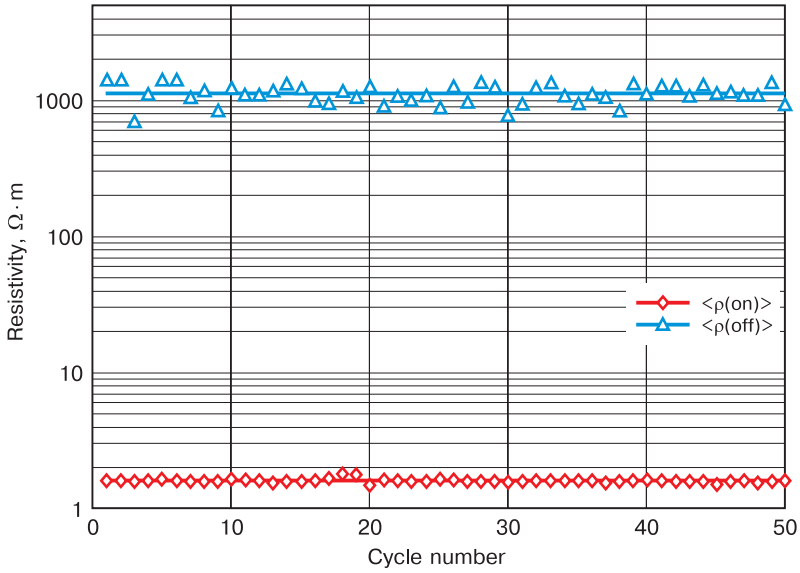



**Рис. 1.18.** Сопротивления в «Выкл» и «Вкл» состояниях при циклическом переключении

## 1.6. МЕМРИСТОРНО-ДИОДНЫЙ КРОССБАР — НОВЫЙ КОМПОНЕНТ НАНОЭЛЕКТРОНИКИ КАК ОСНОВА АППАРАТНОГО УСТРОЙСТВА БИОМОРФНОГО НЕЙРОПРОЦЕССОРА

В настоящем параграфе представлены новые компоненты наноэлектроники — мемристорно-диодные кроссбары, необходимые для создания сверхбольших запоминающей и логических матриц [39] биоморфного нейропроцессора [40].

### 1.6.1. Мемристорно-диодный кроссбар для запоминающей матрицы

В работе [33] для цифровой запоминающей матрицы уже применялся кроссбар с ячейками из последовательно включенных мемристора и обычного диода, требующими высокой идентичности характеристик.

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 низкопроводящего и высокопроводящего состояний мемристоров в паре. Таким образом, в ячейку можно записать три состояния, если мемристоры работают в режиме ключа, и больше трех состояний, если мемристорный материал обеспечивает плавное переключение и несколько устойчивых значений проводимости.

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.

Авторы [33] и другие, высказывая возможность реализации комплементарного соединения мемристоров, не предлагают способов разделения цепей записи и считывания, которые требуются при объединении таких ячеек в сверхбольшие матрицы в запоминающих устройствах и нейропроцессорах. В работе [41] предложено решение этой проблемы путем создания новой ячейки памяти, ее схемотехники и топологии на основе комплементарных мемристоров, соединенных через разделяющие диоды Зенера. Анализируются достоинства и недостатки двух типов конструкций сверхбольших матриц запоминающих устройств — с параллельным выводом данных и с побитным доступом.

Топология комплементарной мемристорно-диодной ячейки, позволяющей организовать матрицу запоминающего устройства с параллельным выводом данных, представлена на рис. 1.13, *б*. На подложке в изолирующем материале создана матрица вертикально ориентированных диодов с эффектом Зенера. Метод изготовления таких диодов по интегральной технологии представлен в работе [42]. Внизу сформированы области легированного и высоколегированного полупроводника *p*-типа, являющиеся катодами диодов Зенера. Катоды 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 *p*-типа, располагается под мемристивным слоем и представляет собой общий контакт комплементарных мемристоров в соответствии с электрической схемой (рис. 1.19, *а*). Комплементарные мемристоры образованы внутри активного слоя между анодным контактом диода и двумя верхними проводниками, которые показаны сверху уходящими проводниками на топологическом чертеже (рис. 1.19, *б*).

Другой тип топологии комплементарной мемристорно-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. 1.19, *в*, 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, являющийся общей шиной последовательного вывода данных. Роль катодного слоя может играть подложка легированного полупроводника *p*-типа.

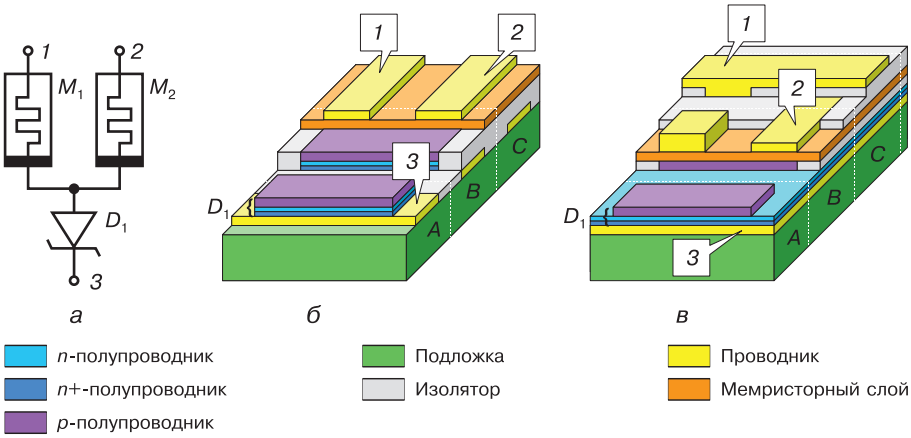



**Рис. 1.19.** Схема и топология ячеек:

*а* — схема комплементарной мемристорно-диодной ячейки;
*б* — топология ячейки для матрицы с параллельным выводом информации;
*в* — топология ячейки для матрицы с последовательным выводом информации

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.

Предлагаемый новый компонент электроники — комбинированный кроссбар [43] (см. рис. 1.19, *б*, 1.20), необходимый для создания сверхбольшой 3D запоминающей матрицы с параллельным выводом данных [44], образован из ячеек, содержащих два мемристора $M_1$ и $M_2$ с общим электродом, совмещенным с анодом диода Зенера $D_1$. 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*.

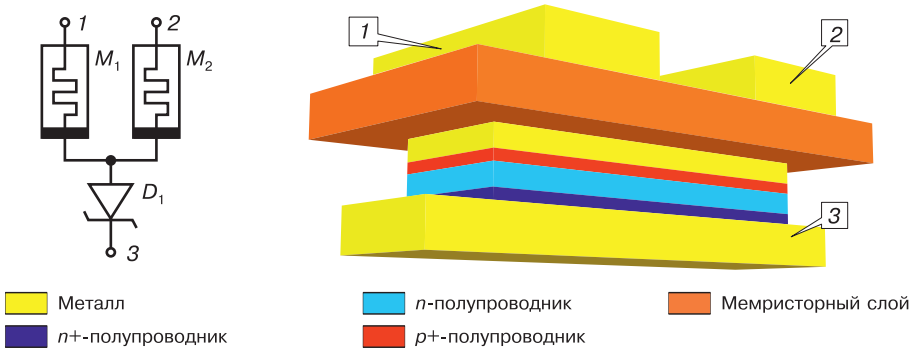



**Рис. 1.20.** Электрическая схема
и топология отдельной ячейки кроссбара

Мемристорный слой и полупроводниковые слои диода изготавливаются промышленным способом — 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 [45].

### 1.6.2. Мемристорно-диодный кроссбар для логической матрицы

Электрическая схема элементарной ячейки логической матрицы [39], показанная на рис. 1.21, представляет собой объединение мемристоров с селективными диодами Зенера, подключенными к одному из проводников кроссбара. В свою очередь этот проводник соединен с затвором КМОП-инвертора. Ячейка имеет несколько входов $I_1$–$I_3$, подключенных к затворам полевых транзисторов $T_1$ и $T_2$ через диоды Зенера $D_1$–$D_3$, и мемристоры $M_1$–$M_3$.

Мемристоры подключены к соединенным затворам полевых транзисторов $T_1$ и $T_2$, включенных комплементарно по схеме КМОП-инвертора. Вход инвертора соединен с проводником $P_1$, выходящим на периферию блока, при этом каждая ячейка имеет свой, не подключенный к другим ячейкам проводник, который является цепью программирования мемристоров. Рядом расположенные элементарные ячейки одного уровня соединены с шинами питания $V_{dd}$ и $V_{ss}$, через которые осуществляется управление режимами работы блока. Напряжением питания управляют драйвера, подключенные к входным шинам, которые вынесены на периферию блока.

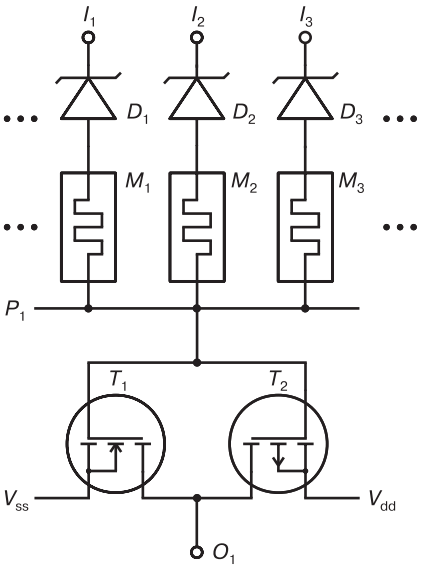



**Рис. 1.21.** Электрическая схема элементарной ячейки для многослойного логического блока

На рис. 1.21 троеточием показано, что мемристорных входов может быть много. Конкретное количество мемристоров зависит от электрических свойств и размеров элементов ячейки. Мемристоры следующих ячеек, подключенных на выход инвертора $O_1$, могут находиться в проводящем состоянии в соответствии с его нагрузочной способностью. В основном режиме работы элементарная ячейка питается по линиям $V_{dd}$ и $V_{ss}$ напряжением ниже напряжений туннельного пробоя диодов Зенера и порога переключения мемристора. При низком напряжении питания исключены изменения сопротивлений мемристоров в логическом блоке, и мемристоры находятся в режиме хранения своих сопротивлений. В режиме программирования подается напряжение питания на шины $V_{dd}$ и $V_{ss}$ больше напряжений туннельного пробоя диодов Зенера и порога переключения мемристоров, при этом осуществляется программирование блока.

Комбинированный кроссбар, состоящий из мемристоров с диодами Зенера, предлагается изготавливать с помощью вакуумной магнетронной технологии. Она заключается в нанесении мемристорного слоя и легированных полупроводниковых слоев диода Зенера, как и проводящих дорожек, в магнетронном технологическом модуле с двумя одновременно распыляющимися катодами.

Сначала на подложку через маску наносятся нижние проводники кроссбара $P_1$–$P_8$, а также наращиваются $V_{ss}$, $V_{dd}$ путем распыления металлического катода. Затем пустое пространство между проводниками методом реактивного магнетронного распыления заполняется изолятором (например, диоксидом кремния). Этот слой изолирует питающие шины. Затем реактивным магнетронным распылением наносится пленка оксида переходного металла (например, $TiO_2$), являющаяся мемристорным слоем. Далее через маску последовательно формируются три полупроводниковых слоя диодов Зенера с разной проводимостью в результате одновременного распыления в магнетроне катодов кремния и легирующей примеси.

## 1.7. СПЕЦИАЛИЗИРОВАННАЯ ПРОГРАММА MDC-SPICE ДЛЯ РАСЧЕТА БОЛЬШИХ ЭЛЕКТРИЧЕСКИХ СХЕМ, СОДЕРЖАЩИХ МЕМРИСТОРНО-ДИОДНЫЕ КРОССБАРЫ

Специализированная программа MDC-SPICE разработана для расчета больших электрических схем, содержащих мемристорно-диодные кроссбары. Этот симулятор является модифицированной версией LTSPICE фирмы Linear Technology подобного симулятора NGSPICE. Краткое описание симулятора представлено в работе [46], с помощью которой во входном устройстве нейропроцессора на основе логической матрицы проведено численное моделирование процесса кодирования информации в биоморфные импульсы.

Электрическая схема в классической программе SPICE задается как набор элементов электроники и связей между ними. 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. Алгоритм расчета переходного процесса является многошаговым. 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.

Набор элементов электроники для электрической схемы программы SPICE включает SPICE-модели транзисторов, диодов и т.д. Транзисторные модели использовались из встроенной библиотеки LT SPICE pmos4 и nmos4. Уровень модели Level=3 допускает установку основных параметров, описание которых для современных полевых транзисторов представлено в [47].

В [44] построена SPICE-модель элементарной ячейки нового компонента электроники — комбинированного мемристорного кроссбара с диодом Зенера. В программу MDC-SPICE, предназначенную для расчета больших схем, добавлена модель мемристора [48], в которой изменения параметра состояния дополнительно были жестко зафиксированы в интервале от 0 до 1. Такое ограничение необходимо, поскольку неабсолютная точность рациональных чисел в компьютерной системе приводит к выходу параметра состояния за границы допустимого интервала и, как следствие,

к неправильной работе модели. Кроме этого, с целью ускорения расчета нелинейная вольт-амперная характеристика, используемая в классической идеализированной SPICE-модели диода Зенера (.model Zener D $R_{on} = R_{off} = 1$ G $V_{fwd} = 0{,}3$ $V_{rev} = 3$), была упрощена и представлена кусочной функцией из трех прямых линий. Таким образом, симулятор заменяет диод резистором с соответствующим значением сопротивления. Вносимая при этом упрощении ошибка мала, когда мемристорно-диодный кроссбар работает в цифровом режиме, и напряжение на диодах Зенера не приближается к пороговым значениям открытия и обратного пробоя диода.

Выбор в пользу этой модели диода сделан на основании того, что существующая детальная макроскопическая модель [49] при совпадении задействованных в моделировании участков вольт-амперных характеристик (рис. 1.22, *а*) обладает существенным недостатком, связанным с неправильным моделированием сопротивления утечки (рис. 1.22, *б*). Кроме этого, расчет матрицы из большого числа ячеек с использованием идеализированной модели требует намного меньше времени.

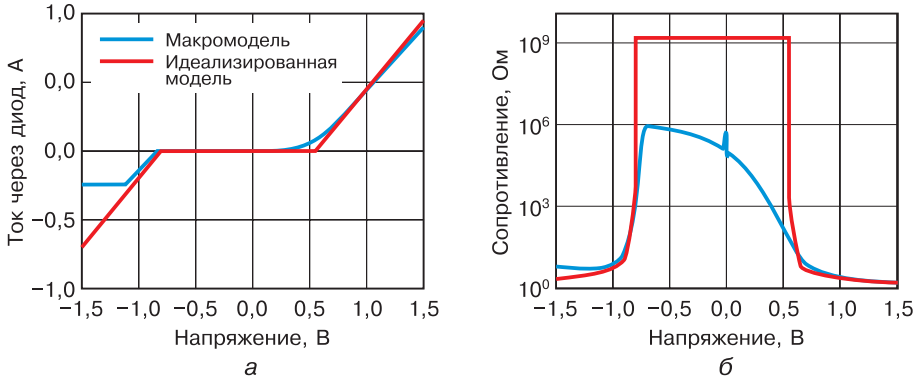



**Рис. 1.22.** Сравнение идеализированной и детальной моделей диода Зенера:

*а* — вольт-амперные характеристики;
*б* — сопротивление диода Зенера от напряжения на его контактах

Для подачи напряжений в матрицу были использованы управляемые источники напряжения с заданной последовательностью изменения потенциала на выходе, имитирующие работу периферийных коммутационных схем на полевых транзисторах.

Разработанная специализированная программа MDC-SPICE была использована для моделирования процесса записи в запоминающей матрице на основе ячейки из комплементарных мемристоров с диодом и без него.

Процесс записи в запоминающую ячейку из комплементарных мемристоров и диода Зенера происходит в два этапа. Сначала на контакты *1* и *3* (см. рис. 1.20) подаются потенциалы разного знака, которые по абсолютной

величине меньше порога изменения состояния мемристора. Потенциалы на контактах *2* и *3* при этом равны. Итоговое напряжение между точками *1* и *3* оказывается выше порогового, а между точками *2* и *3* равно нулю. Таким образом, происходит изменение состояния только одного мемристора в ячейке. Второй этап записи заключается в изменении состояния второго мемристора в ячейке и выполняется аналогично первому.

Для исключения паразитной записи в матрице без диодов [50] на незадействованные шины подается половина напряжения записи, при этом напряжение на контактах невыбранных мемристоров остается ниже порогового. Диод Зенера позволяет подавать напряжения только на выбранную ячейку, выступая в качестве селективного элемента и предотвращая паразитную запись в соседние ячейки кроссбара через смежные шины.

Результаты моделирования показывают, что поддержание потенциала на шинах незадействованных ячеек вносит значительный вклад в итоговое потребление запоминающей матрицы. Как видно из рис. 1.23, затраты энергии на запись одной ячейки из комплементарных мемристоров в матрице размером 100 × 100 снижаются в 8 раз при добавлении в каждую ячейки диода Зенера. В обоих вариантах использовались худшие условия для записи, при которых все ячейки матрицы изначально находились в одинаковом состоянии. Максимальное и минимальное сопротивления мемристоров равно, соответственно, 110 и 10 кОм.

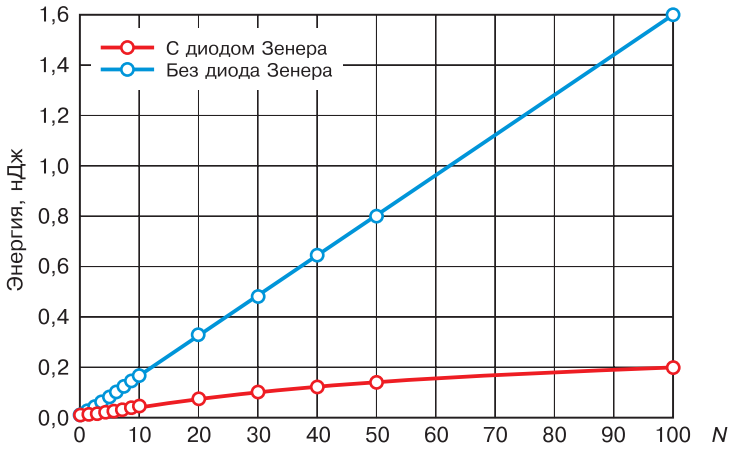



**Рис. 1.23.** Затраты энергии на запись одной ячейки комплементарных мемристоров без диода и с диодом Зенера в зависимости от числа ячеек в квадратной матрице $N \times N$

Как показывает результат моделирования, включение нелинейного селективного элемента — диода Зенера в комплементарную ячейку запоминающей матрицы [44] обеспечивает минимизацию паразитных токов при записи и считывании. Уменьшение паразитных токов при записи и отсутствие

необходимости поддерживать на незадействованных шинах напряжения приводит к снижению на порядок затрат на запись одной ячейки по сравнению с матрицей без диодов, предложенной в [50].

## 1.8. КОНЦЕПЦИЯ АППАРАТНОГО УСТРОЙСТВА БИОМОРФНОГО НЕЙРОПРОЦЕССОРА

Под биоморфным нейропроцессором авторы подразумевают аппаратное средство, представляющее собой программно-аппаратную нейросеть, построенную на основе биологической электрической модели нейрона Ходжкина–Хаксли [51; 52]. Программная часть такой реализации отвечает за возможность программирования синаптических связей между нейронами нейросети, а также ввод и вывод информации. Все основные функции: взвешивание и суммирование импульсов напряжения (умножение на вес связи), генерация потенциала действия при превышении порога, маршрутизация импульсов между нейронами выполняются с помощью аппаратного средства. С этой точки зрения, такая реализация ближе к полностью аппаратной.

В концепции аппаратной реализации нейропроцессора предлагаются два подхода [53]. Первый подход направлен на уменьшение числа элементов электроники при использовании аналоговых вычислений для синапсов и нейронов. Схема нейропроцессора в этом случае предполагает наличие запоминающей матрицы, блока нейронов и маршрутизатора на основе логической матрицы. Аналоговая запоминающая матрица является массивом синапсов и, помимо запоминания информации, производит часть расчетов нейросети в виде взвешенного суммирования входных импульсов нейронов. Нейронный блок производит оставшуюся часть вычислений, относящихся к процессам зарядки мембраны нейрона выходными импульсами запоминающей матрицы и генерации выходных импульсов нейронов при превышении порога активации. Маршрутизатор отвечает за перенаправление выходных сигналов нейронов на синапсы в запоминающей матрице.

Входное устройство нейропроцессора предназначено для первичной обработки звуковой и видеоинформации путем ее сжатия и кодирования в виде отдельных импульсов, в том числе подобных биоморфным импульсам мозга. Выходное устройство осуществляет преобразование информации об активации нейронов в цифровой двоичный код и передачу на интерфейсный блок.

Второй подход основан на унификации электронных компонентов за счет использования электрической схемы логической матрицы во всех функциональных блоках нейропроцессора. Универсальная электрическая схема логической матрицы может выполнять расчет связей нейросети в отсутствие отдельной запоминающей матрицы. На основе собственных

логических функций она выполняет умножение матрицы на вектор путем последовательных конъюнкций с инверсией; в качестве маршрутизатора направляет выходные импульсы нейронов на синапсы других нейронов; в качестве части входного устройства нейропроцессора выполняет первичную обработку сигнала в цифровом режиме с помощью умножения матрицы на вектор, преобразуя входные данные в нужный формат; в качестве части выходного устройства осуществляет сжатие информации с помощью того же умножения для передачи в интерфейсный блок.

Два варианта функциональной схемы нейропроцессора, в которой отражены основные узлы, представлены на рис. 1.24.

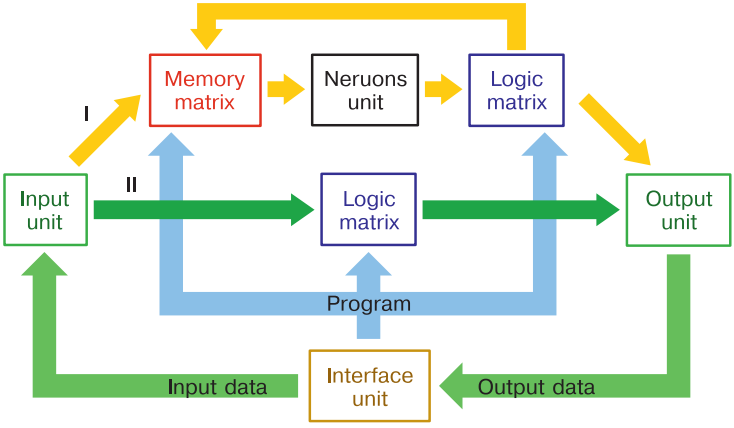



**Рис. 1.24.** Функциональная схема нейропроцессора:
I — с использованием запоминающей матрицы для синапсов и логической матрицы в качестве маршрутизатора; II — на основе универсальной логической матрицы в отсутствие запоминающей матрицы

Достоинства реализации нейропроцессора на основе второго подхода проявляются в более высоком быстродействии и энергоэффективности, обусловленные представлением сигналов в цифровой форме, и более простой технологии изготовления аппаратной базы нейропроцессора за счет использования унифицированных элементов во всех блоках.

Биоморфный нейропроцессор, способный воспроизводить работу кортикальной колонки, должен обеспечить вычисление большого числа нейронов. Соответственно, большое количество элементов в электрической схеме нейропроцессора, построенного на основе предлагаемой концепции, накладывает на его узлы повышенные требования: высокую степень интеграции элементов при объединении их в сверхбольшую матрицу; минимизацию площади, которую занимает ячейка матрицы на кристалле; высокие быстродействие и энергоэффективность; простую технологию изготовления.

## 1.9. АППАРАТНАЯ РЕАЛИЗАЦИЯ НЕЙРОПРОЦЕССОРА

### 1.9.1. Запоминающая матрица как массив мемристорных синапсов, задающий вес связи между нейронами

В варианте I нейропроцессора (см. рис. 1.24) один синапс нейрона аппаратно представлен ячейкой запоминающей матрицы. Организация связи нейрона через синапсы со всеми остальными нейронами в нейронном блоке осуществляется матричным соединением ячеек через шины. Не существует подходящих матриц для применения в качестве сверхбольшого массива синапсов для биоморфного нейропроцессора.

Массивы мемристорных ячеек типа 1T1R [54; 55] с селективным транзистором используются при построении аппаратных нейросетей [56; 57]. Для аппаратной реализации импульсной нейросети предложена ячейка типа 2T1R [58], в которой второй транзистор используется в механизме самообучения. В работе [59] предложена ячейка типа 1T1R с новым типом транзистора, который обладает малым размером и может менять знак основных носителей заряда, что позволяет свести 2T1R ячейки на обычных транзисторах к ячейке типа 1T1R. Однако такие ячейки обладают сравнительно низкой интеграцией из-за наличия дополнительных управляющих проводников и большой площади транзисторов по сравнению с размерами мемристоров. Состоящие только из мемристоров ячейки 4M1M [60] и 2M1M [61] позволяют производить логические операции и хранить двоичные данный, но не предназначены для использования в аналоговых вычислениях.

С точки зрения масштабирования сверхбольшой запоминающей матрицы необходим выбор селективного элемента малого размера. Ячейки типа 1S1R [62; 63] и 1D1R являются наиболее оптимальными с точки зрения занимаемой площади и нелинейности ВАХ [62]. Среди селективных элементов выбор сделан в пользу диода Зенера по следующим причинам: низкое напряжение пробоя, используемое для перепрограммирования мемристоров, как унифицированный элемент он же используется в логической матрице для реализации диодной логики. Селективные элементы, предложенные в [62; 63], обладают симметричной вольт-амперной характеристикой, что не позволяет их использовать для реализации логических операций. Соответственно, в запоминающей матрице тоже не желательно их использовать, поскольку это приведет к дополнительным шагам при производстве нейропроцессора.

В работе [44] разработана электрическая схема, топология и нанотехнология изготовления сверхбольшой многослойной запоминающей матрицы с энергонезависимой памятью и высокой степенью интеграции элементов на основе комбинированного мемристорно-диодного кроссбара.

В подтверждение высокой степени интеграции приведем табл. 1.1, в которой даны площади одной ячейки на основе мемристоров в представленной и известных запоминающих матрицах.

*Таблица 1.1*

**Площадь ячейки при различном включении мемристоров**

| Структура синапса | Площадь $F^2$ | Ссылка |
|---|---|---|
| 2M | 8 | [35] |
| 1D2M | 8 | Presented work |
| 2 × 1T1M | 24,8 | [40] |
| 2T1M | 40,3 | [38] |

Площадь ячейки матриц [58; 54] определяется суммарной площадью транзисторов, поскольку предполагается, что мемристор расположен над ними. Для оценки был использован размер планарных КМОП-транзисторов из работы [59].

### 1.9.2. Развитие электрической модели нейрона для интеграции запоминающей матрицы с блоком нейронов

Электрические схемы всех нейронов нейронного блока одинаковы. За основу электрической схемы нейрона взята известная биоморфная электрическая модель нейрона [51; 52]. Функциональная схема, показанная на рис. 1.25, объединяет два нейрона из нейронного блока, выходы которых скоммутированы через логическую матрицу на входы синапсов третьего нейрона Input 1 и Input 2 этого же блока.

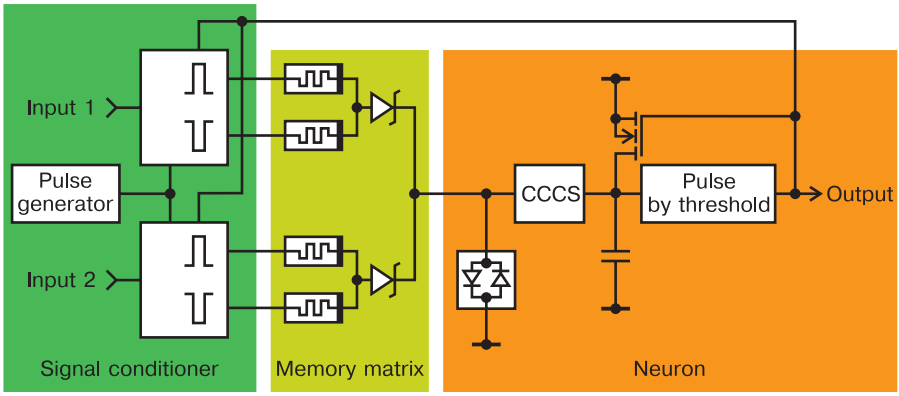



**Рис. 1.25.** Развитая электрическая схема нейрона с двумя синапсами (в запоминающей матрице), которые соответствуют входам от двух других нейронов

Развитие электрической модели нейрона заключается в изменении цепи формирования импульса потенциала действия, добавлении отдельной цепи разряда конденсатора и нелинейного элемента, предназначенного для перепрограммирования мемристоров. Суммарный ток синапсов поступает на вход ИТУТ (источника тока, управляемый током). ИТУТ заряжает конденсатор током, который пропорционален сумме выходных токов ячеек. При превышении напряжения на конденсаторе заданного порога в специальном блоке происходит генерация одиночного импульса, являющегося выходным импульсом нейрона. Этот же импульс управляет сбросом заряда на конденсаторе и увеличивает амплитуды напряжений входных импульсов выше порога программирования мемристоров и порога срабатывания нелинейного элемента в нейроне. Таким образом, при наличии выходного сигнала с нейрона и входного сигнала на синапсе происходит увеличение проводимости одного из мемристоров комплементарной пары и, соответственно, усиление синаптической связи.

### 1.9.3. Логическая матрица как массив мемристорных синапсов, задающий маршрут связи между нейронами

Не существует подходящих матриц логики для применения в качестве маршрутизатора для биоморфного нейропроцессора. Известный массив Акерса на основе мемристоров [64] можно запрограммировать на выполнение любой логической функции. Однако в одном массиве невозможно реализовать основную логическую функцию — комбинационную схему умножения вектора на матрицу из-за наличия в нем только одного выхода. Использование нескольких массивов, необходимое для организации такой операции, приведет к росту количества элементов и, соответственно, к увеличению размеров логического устройства. При этом мемристорный массив Акерса обладает слабой интеграцией элементов, связанной с большим количеством транзисторов в ячейке, и высокой деградацией выходного сигнала при большом размере матрицы. Если же использовать массив Акерса в последовательностной (*sequential*) схеме, это приведет к увеличению времени вычислений из-за последовательного вычисления каждого разряда выходного вектора.

В отличие от массива Акерса устройство Hewlett-Packard (HP) [54] выполняет умножение матрицы чисел на вектор в аналоговом виде. Оно может быть использовано в качестве цифрового логического блока при подаче входных логических сигналов на затворы транзисторов. Но использование такой большой матрицы (8192 ячеек) в качестве сверхбольшой логической матрицы нейропроцессора нецелесообразно из-за низкой интеграции элементов: на один транзистор с минимальным размером $4F^2$ приходится лишь один мемристор размером $1F^2$.

В [39; 40] разработаны электрическая схема, топология и технология изготовления сверхбольшой 3D логической матрицы (более $10^6$ ячеек) на основе комбинированного мемристорно-диодного кроссбара 1D1M- и КМОП- (CMOS) инверторов. Над КМОП-инвертором минимального размера ~$8F^2$ можно разместить до 9 мемристоров. Дальнейшее увеличение числа мемристоров в кроссбаре потребует применения транзисторов большего размера. Преимуществом предложенной конструкции 3D-матрицы из многовходовых элементов И–НЕ по сравнению с массивами [64] и [54] является более высокая степень интеграции, которая достигается за счет объединения на кристалле в 3D-структуру одинаковых перпендикулярно ориентированных функциональных пластов с комбинированными мемристорными кроссбарами в качестве коммутаторов и, как следствие, компактного расположения друг над другом элементов ячейки и самих ячеек.

Для оценки степени интеграции приведем удельную площадь, полученную делением общей площади матрицы на число мемристоров в ней $N_{mem}$, и отношение числа мемристоров $N_{mem}$ к числу транзисторов $N_{tran}$ в логической матрице (табл. 1.2) при использовании планарной топологии транзисторов, приведенной в работе [59]. Планарная геометрия транзисторов, представленная в [59], позволяет разместить над ними до 8 мемристоров размером $1F^2$.

*Таблица 1.2*

**Сравнение размеров логических матриц**

| **Тип ячейки** | **Удельная площадь, $F^2$** | $N_{mem}/N_{tran}$ | **Ссылка** |
|---|---|---|---|
| 4T1M (Акерс) | 49,6 | 0.25 | [64] |
| 1T1M (HP) | 12,4 | 1 | [54] |
| 2T8M | 3,1 | 4 | Presented work |

Из таблицы видно, что использование 3D-интеграции для расположения мемристоров над транзисторами позволяет значительно уменьшить площадь логической матрицы.

Логическая матрица, кроме собственно логических операций, способна выполнять обработку оцифрованных выходных импульсов нейронов из блока нейронов и их маршрутизацию на синапсы запоминающей матрицы, соединенных с другими нейронами из этого же блока.

В матрице используется одна логическая функция И–НЕ, поскольку набор таких элементов достаточен для осуществления маршрутизации импульсов. Добавление отдельной функции инверсии, с одной стороны, повышает гибкость системы за счет образования полного логического базиса, но с другой стороны приведет к дополнительному увеличению элементов в сверхбольшой матрице

Список литературы


1. *Pickett M.D.*, *Strukov D.B.*, *Borghetti J.L.* et al. Switching dynamics in titanium dioxide memristive devices // Journal of Applied Physics. 2009. Vol. 106. P. 074508.
2. *Simmons J.G.* Electric tunnel effect between dissimilar electrodes separated by a thin insulating film // Journal of Applied Physics. 1963. Vol. 34. P. 2581.
3. *Abdalla H.*, *Pickett M.D.* SPICE modeling of memristors // IEEE International Symposium of Circuits and Systems (ISCAS). 2011. Pp. 1832–1835.
4. *Журавский Д.В.*, *Бобылев А.Н.*, *Удовиченко С.Ю.*, *Филиппов В.А.* Установление подобия свойств синапса и мемристора, используемого в электронном устройстве // Нейрокомпьютеры: разработка и применение. 2015. № 11. С. 95–101.
5. *Бобылев А.Н.*, *Удовиченко С.Ю.* Создание электронного запоминающего устройства, подобного по свойствам синапсу мозга // Доклады Томского государственного университета систем управления и радиоэлектроники. 2015. № 4 (38). С. 68–71.
6. *Bobylev 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. Vol. 45. No. 6. Pp. 396–401.
7. *June S.K.*, *Young H.D.*, *Yoon Ch.B.* et al. Roles of interfacial $TiO_x N_{1-x}$ layer and TiN electrode on bipolar resistive switching in TiN/$TiO_2$/TiN frameworks // Applied Physics Lettters. 2010. Vol. 96. P. 223502.
8. *Chang T.*, *Jo S.-H.*, *Lu W.* Short-term memory to long-term memory transition in a nanoscale memristor // ACS Nano. 2011. Vol. 5. No. 9. Pp. 7669–7676.
9. *Jo S.H.*, *Chang T.*, *Ebong I.* et al. Nanoscale memristor device as synapse in neuromorphic systems // Nano Letters. 2010. Vol. 10. No. 4. Pp. 1297–1301.
10. *Миркес Е.Н.* Нейрокомпьютер. Проект стандарта // Новосибирск: Наука, Сибирская издательская фирма РАН. 1998. 337 с.
11. *Govoreanu B.*, *Redolfi A.*, *Zhang L.* et al. Vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM): An area-scalable switching current, self-compliant, highly nonlinear and wide on/off-window resistive switching cell // IEEE International Electron Devices Meeting. 2013. Pp. 10.2.1–10.2.4.
12. *Strukov D.B.*, *Snider G.S.*, *Stewart D.R.*, *Williams R.S.* The missing memristor found // Nature. 2008. Vol. 453. Pp. 80–83.
13. *Strukov D.B.*, *Williams R.S.* Exponential ionic drift: Fast switching and low volatility of thin-film memristors // Applied Physics A. 2009. Vol. 94. Pp. 515–519.
14. *Rozenberg M.J.*, *Sanchez M.J.*, *Weht R.* et al. Mechanism for bipolar resistive switching in transition-metal oxides // Physical Revier. B. 2010. Vol. 81. 115101.
15. *Dirkmann S.*, *Kaiser J.*, *Wenger C.*, *Mussenbrock T.* Filament growth and resistive switching in hafnium oxide memristive devices // ACS Applied Materials and Interfaces. 2018. Vol. 10. No. 17. Pp. 14857–14868.
16. *Vandelli L.*, *Padovani A.*, *Larcher L.* et al. A physical model of the temperature dependence of the current through $SiO_2$/$HfO_2$ stacks // IEEE Transactions on Electron Devices. 2011. Vol. 58. No. 9. Pp. 2878–2887.
17. *Chernov A.A.*, *Islamov D.R.*, *Piknik A.A.* et al. Three-dimensional non-linear complex model of dynamic memristor switching // ECS Transactions. 2017. Vol. 75. No. 32. Pp. 95–104.
18. *Бусыгин А.Н.*, *Ибрагим А.Х.*, *Удовиченко С.Ю.* Математическое моделирование резистивных состояний и динамического переключения мемристора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2020. Т. 6. № 2. С. 127–144.

19. *Френкель Я.И.* К теории электрического пробоя в диэлектриках и электронных полупроводниках // Журнал экспериментальной и теоретической физики. 1938. Т. 8. № 12. С. 1292–1301.
20. *Hill R.M.* Poole-Frenkel conduction in amorphous solids // Philosophical Magazine. 1971. Vol. 23. Pp. 59–86.
21. *Matveyev Y.*, *Kirtaev R.*, *Fetisova A.* et al. Crossbar nanoscale $HfO_2$ — based electronic synapses // Nanoscale research letters. 2016. Vol. 11. Pp. 147.
22. *Menzel S.*, *Salinga M.*, *Bцttger U.*, *Wimmer M.* Physics of the switching kinetics in resistive memories // Advanced Functional Materials. 2015. Vol. 25. Pp. 6306–6325.
23. *Исламов Д.Р.*, *Грищенко В.А.*, *Чин А.* О транспорте заряда в тонких пленках оксида гафния и циркония // Автометия. 2017. Т. 53. № 2. С. 102–108.
24. *Ielmini D.*, *Waser R.* Resistive switching. from fundamentals of nanoionic redox processes to memristive device applications // Wiley-VCH.Germany. 2016. 784 p.
25. *Горшков О.Н.*, *Антонов И.Н.*, *Белов А.И.* и др. Изучение диффузии ионов кислорода в МДМ-структурах на основе стабилизированного диоксида циркония, проявляющих резистивное переключение // Вестник Нижегородского университет им. Н.И. Лобачевского. 2013. № 5 (1). С. 51–54.
26. *Savelev S.E.*, *Alexandrov A.S.*, *Bratkovsky A.M.*, *Williams R.S.* Molecular dynamics simulations of oxide memory resistors (memristors) // Nanotechnology. 2011. Vol. 22. P. 254011.
27. *Kumar S.* et al Oxygen migration during resistance switching and failure of hafnium oxide memristors // Applied Physics Lettters. 2017. Vol. 110. P. 103503.
28. *Noman M.* et al Computational investigations into the operating window for memristive devices based on homogeneous ionic motion // Applied Physics A. 2011. Vol. 102. Pp. 877–883.
29. *Walczyk C.*, *Walczyk D.*, *Schroeder T.* Impact of Temperature on the Resistive Switching Behavior of Embedded $HfO_2$-Based RRAM Devices // IEEE Transactions on Electron Devices. 2011. Vol. 58. No. 9. Pp. 3124–3131.
30. *Merolla P.A.* et al. A million spiking-neuron integrated circuit with a scalable communication network and interface // Science. 2014. Vol. 345. Pp. 668–672.
31. *Prezioso M.*, *Merrikh-Bayat F.*, *Hoskins B.D.* et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors // Nature. 2015. Vol. 521. Pp. 61–64.
32. *Kim K.-H.*, *Gaba S.*, *Wheeler D.* et al. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications // Nano Letters. 2012. Vol. 12. Pp. 389–395.
33. *Liu T.* et al. A 130.7-mm 2-layer 32-Gb ReRAM memory device in 24-nm technology // IEEE J. Solid-State Circuits. 2014. Vol. 49. Pp. 140–153.
34. *Jo S.H.*, *Chang T.*, *Ebong I.* et al. Nanoscale memristor device as synapse in neuromorphic systems // Nano Letters. 2010. Vol. 10. Pp. 1297–1301.
35. *Bobylev A.N.*, *Busygin A.N.*, *Pisarev A.D.* et al. Neuromorphic coprocessor prototype based on mixed metal oxide memristors // International Journal of Nanotechnology. 2017. Vol. 14. No. 7/8. Pp. 698–704.
36. *June S.K.*, *Young H.D.*, *Yoon Ch.B.* et.al. Roles of interfacial $TiO_xN_{1-x}$ layer and TiN electrode on bipolar resistive switching in $TiN/TiO_2/TiN$ frameworks // Applied Physics Lettters. 2010. Vol. 96. P. 223502.
37. *Алехин А.П.*, *Батурин А.С.*, *Григал И.П.* и др. Мемристор на основе смешанного оксида металлов // 2013. Патент № 2472254 РФ. Патентообладатель МФТИ.
38. *Jo S.H.*, *Chang T.*, *Ebong I.* et al. Nanoscale memristor device as synapse in neuromorphic systems // Nano Letters. 2010. Vol. 10. No. 4. Pp. 1297–1301.

39. *Udovichenko S.*, *Pisarev A.*, *Busygin A.*, *Maevsky O.* 3D CMOS, memristor nanotechnology for creating logical and memory matrics of neuroprocessor // Nanoindustry. 2017. No. 5. Pp. 26–34.
40. *Udovichenko S.Yu.*, *Pisarev A.D.*, *Busygin A.N.*, *Maevsky O.V.* Neuroprocessor based on combined memristor-diode crossbar // Nanoindustry. 2018. No. 5. Pp. 344–355.
41. *Maevsky O.V.*, *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.* Complementary memristor-diode cell for a memory matrix in neuromorphic processor // International journal of nanotechnology. 2018. Vol. 15. No. 4/5. Pp. 388–393.
42. *Weize Chen*, *Xin Lin*, *Parris P.M.* Zener diode devices and related fabrication methods // 2017. Patent US 9397230 B2.
43. *Писарев А.Д.*, *Бусыгин А.Н.*, *Бобылев А.Н.*, *Удовиченко С.Ю.* Комбинированный мемристорно-диодный кроссбар как основа запоминающего устройства // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2017. № 4. С. 142–149.
44. *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. Vol. 198. Pp. 1–7.
45. *Kim H.K.*, *Li C.C.*, *Fang X.M.*, *Solomon J.* et al. Erbium doped semiconductor thin films prepared by RF magnetron sputtering // Materials Research Society Symposia Proceedings. 1993. Vol. 301. Pp. 55–60.
46. *Ибрагим А.Х.*, *Удовиченко С.Ю.* Моделирование устройства кодирования информации для импульсной аппаратной нейросети // Сб. научных трудов «Математическое и информционное моделирование». 2020. Вып. 18. С. 10–16.
47. *Vladimirescu A.*, *Liu S.* The Simulation of MOS Integrated Circuits Using SPICE2 // Univ. California, ERL Memo. ERL-M80/7. 1980, 86 p.
48. *Biolek D.*, *Di Ventra M.*, *Pershin Y.V.* Reliable SPICE simulations of memristors, memcapacitors and meminductors // Radioengineering. 2013. Vol. 22. No. 4. Pp. 945–968.
49. *Wong S.*, *Hu C.M.* SPICE macro model for the simulation of Zener diode I–V characteristics // IEEE Circuits and Devices Magazine. 1991. Vol. 7. No. 4. Pp. 9–12.
50. *Zhao W.*, *Portal M.*, *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. Vol. 74. No. 6. Pp. 2484–2496.
51. *Brette R.*, *Gerstner W.* Adaptive exponential integrate-and-fire model as an effective description of neuronal activity // Journal of Neurophysiology. 2005. Vol. 94. Pp. 3637–3642.
52. *Hodgkin A.L.*, *Huxley A.F.* A quantitative description of membrane current and its application to conduction and excitation in nerve // Journal of Physiology. 1952. Vol. 117. No. 4. Pp. 500–544.
53. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor-diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
54. *Li C.*, *Hu M.*, *Li Y.* et al, Analogue signal and image processing with large memristor crossbars // Nature electronics. 2018. Vol. 1. No. 1. Pp. 52–59.
55. *Ghenzi N.*, *Rozenberg M.*, *Pietrobon L.* et al. One-transistor one-resistor (1T1R) cell for large-area electronics // Applied Physics Letters. 2018. Vol. 113. P. 072108.
56. *Yao P.*, *Wu H.*, *Gao B.* et al. Online training on RRAM based neuromorphic network: Experimental demonstration and operation scheme optimization // 2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM). 2017. Pp. 182–183.
57. *Li C.*, *Belkin D.*, *Li Y.* et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks // Nature Communications. 2018. Vol. 9. P. 2385.

58. *Ielmini D.* Brain-inspired computing with resistive switching memory (RRAM): Devices, synapses and neural networks // Microelectronic Engineering. 2018. Vol. 190. Pp. 44–53.
59. *Levisse A.*, *Gaillardon P.E.*, *Giraud B.* et al. Resistive switching memory architecture based on polarity controllable selectors // IEEE Transactions On Nanotechnology. 2019. Vol. 18. Pp. 183–194.
60. *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. Vol. 62. No. 5. Pp. 1402–1411.
61. *Teimoori M.*, *Amirsoleimani A.*, *Ahmadi A.*, *Ahmadi M.* A 2M1M crossbar architecture: Memory // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2018. Vol. 26. No. 12. Pp. 2608–2618.
62. *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. Pp. 31.7.1–31.7.4
63. *Zhang L.*, *Govoreanu B.*, *Redolfi A.* et al. High-drive current (>1 MA/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. Pp. 6.8.1–6.8.4.
64. *Levy Y.*, *Bruck J.*, *Cassuto Y.* et al. Logic operations in memory using a memristive Akers array // Microelectronics Journal. 2014. Vol. 45. Pp. 1429–1437.

# ГЛАВА 2

# БИОМОРФНАЯ НЕЙРОСЕТЬ ДЛЯ НЕЙРОПРОЦЕССОРА

Нейросеть может быть программной, аппаратной или программно-аппаратной (комбинированной). В работе [1] представлена аппаратная реализация LIF (leaky-integrate-fire) нейрона с мемристорными синапсами для целей обнаружения ошибок. Предложенную электрическую схему нейрона можно использовать для реализации аппаратной нейронной сети. Более гибкая архитектура нейросети [2; 3] содержит, помимо связок синапс-нейрон, еще и третий объект — астроцит, обеспечивающий непрямую обратную связь для синапсов нейронов. Если включить в аппаратную реализацию нейронной сети электрическую схему, которая имитирует работу астроцита, то можно увеличивать проводимость оставшихся мемристорных синапсов при обнаружении поврежденного.

Для обработки информации в сверхбольшой нейросети, предназначенной для нейропроцессора с ограниченными вычислительными ресурсами, необходимо построить модель нейрона максимально упрощенную (с точки зрения времени расчета), но без существенной потери точности. Из таких относительно простых нейронов, описываемых биоморфной моделью, можно строить нейросеть, моделирующую работу кортикоморфной колонки с определенными функциями, подобно кортикальным колонкам мозга. Комбинируя кортикоморфные колонки, в дальнейшем можно реализовать модель коры головного мозга, которая не будет требовать для расчета больших вычислительных мощностей, поскольку программные нейросетевые расчеты будут заменены на аппаратную реализацию.

Под программно-аппаратной реализацией подразумевается работа электронного устройства, производящего вычисления синапсов в аналоговом виде и остальных частей нейрона в цифровом. Первая часть вычислений может выполняться в мемристорном кроссбаре, а вторая — в КМОП-микроконтроллере. Из-за ограниченных вычислительных ресурсов автономного аппаратного средства модель нейрона должна обеспечивать достаточную простоту и эффективность расчета сверхбольшой нейросети.

Как известно, в реальном нейроне формирование следа памяти включает в себя целую совокупность последовательно развивающихся и обусловливающих друг друга явлений, включающих в себя как формирование потенциалов в аксонных окончаниях и дендритах клетки (с определенным временем жизни), так и структурные изменения клетки в результате синтеза новых белков под управлением генетического аппарата нейрона,

запускаемого паттернами входных сигналов. К таким структурным изменениям в первую очередь следует отнести синтез новых рецепторов медиаторов и их встраивание в постсинаптические мембраны [4]. Следуя концепции построения биоморфной модели нейрона в [4], в работе [5] использованы механизмы краткосрочной и долгосрочной памяти, основанные на явлении фасилитации в аксоне и релаксации постсинаптического потенциала дендрита, а также изменении количества рецепторов нейромедиатора.

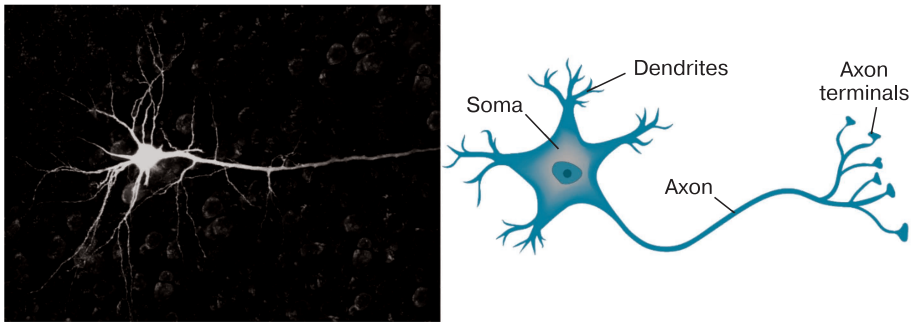
Представленная в [5] оригинальная биоморфная модель нейрона отличается от информационных моделей более сложным устройством нейрона. А от подробных биологических моделей — тем, что в численном решении дифференциального уравнения, описывающего изменение потенциала на мембране нейрона, произведена замена отдельных спайков на среднюю частоту их следования. Замена spiking кодирования информации на кодирование средней частотой потенциалов действия за шаг моделирования позволяет регулировать силу реакции нейрона за один шаг, увеличить шаг по времени и, как следствие, увеличить скорость расчета нейросети путем уменьшения количества шагов. Такой подход применяется при моделировании больших нейросетей [6].

С одной стороны, нас интересуют только процессы, относящиеся к обработке информации, поэтому метаболические реакции, обеспечивающие жизнь клетки, не рассматриваются. С другой стороны, в упрощенной форме учитываются закономерности распространения и генерации сигналов в дендритах и аксонах (в виде простых функций), полученных на основании наблюдений за реальными нейронами.

## 2.1. ОРИГИНАЛЬНАЯ БИОМОРФНАЯ МОДЕЛЬ НЕЙРОНА

Рассматриваемая модель формализует наиболее существенные информационные процессы, происходящие в реальном нейроне при обработке актуальных данных и обучении. Такая модель в функциональном смысле ближе к реальному нейрону (рис. 2.1), чем классические искусственные нейроны.

В этой модели, как и в простой формальной модели, осуществляется сравнение суммарного входного линейного сигнала с некоторым порогом, при превышении которого нейрон активируется и передает возбуждение на выход. Тем не менее, согласно концепции биоморфной модели нейрона [4], связь между нейронами сложнее и состоит из двух частей: аксона пресинаптического нейрона и дендрита постсинаптического нейрона, причем свойства этой связи зависят, как от амплитуды сигнала, так и от его длительности. На рис. 2.2 представлена блок-схема рассматриваемого искусственного биоморфного нейрона.




*а* *б*

**Рис. 2.1.** Фотография с электронного микроскопа (*а*) и отдельные части реального нейрона (*б*)

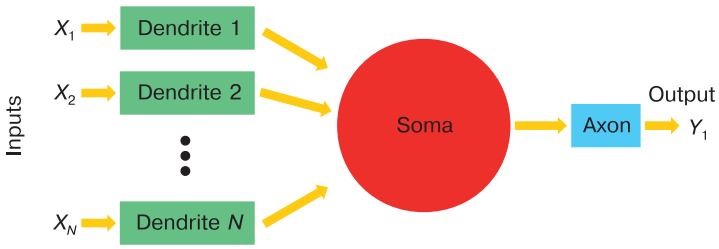



**Рис. 2.2.** Схема биоморфной модели нейрона

Модель нейрона состоит из трех функциональных частей: дендритов, сомы (тела нейрона) и аксона, соответствующих реальным частям нейрона. Дендриты являются входными частями нейрона, аксон — выходной частью. Сома, или тело, нейрона осуществляет сравнение суммы сигналов от дендритов с некоторым пороговым значением. Если порог превышен, то нейрон активируется — переходит в возбужденное состояние. Информация о том, что нейрон возбужден, передается на аксон нейрона и на выходе аксона появляется сигнал определенной формы.

При активации нейрона в нем возникают импульсы потенциала действия, частота которых определяет силу сигнала. Под их действием в аксоне накапливается электрический потенциал (формируется кальциевый микродомен в аксонной терминали), что вызывает выброс молекул нейромедиатора в межсинаптическую щель. Эти молекулы связываются с соответствующими им белковыми рецепторами, расположенными на постсинаптической мембране дендрита другого нейрона. При связывании нейромедиатора с рецептором у последнего изменяется электрический потенциал: если нейромедиатор возбуждающий, то потенциал положительный, тормозящий — отрицательный. Суммарное действие активированных

рецепторов в дендрите называют вызванным постсинаптическим потенциалом (ПСП). Если в соме сумма потенциалов от всех дендритов превышает некий порог активации в аксонном холмике, то нейрон активируется, и сигнал распространяется дальше по аксонной терминали нейрона.

Современная система уравнений, описывающая изменение потенциала во времени на мембране нейрона используется медиками для симуляции воздействия на нейросети в живом мозге [7]. Система включает в себя: дифференциальное уравнение первого порядка для потенциала $V$, в котором присутствует сумма синаптических токов через рецепторы $I(t)$, и уравнение первого порядка для тока адаптации $w$

$$\begin{cases} C\dfrac{\mathrm{d}V}{\mathrm{d}t} = -g_L\left(V - E_L\right) + g_L\Delta_T \exp\left(\dfrac{V - V_T}{\Delta_T}\right) - w + I\left(t\right); & (2.1) \\ \tau_w \dfrac{\mathrm{d}w}{\mathrm{d}t} = \alpha\left(V - E_L\right) - w, & (2.2) \end{cases}$$

где $C$ — электрическая емкость мембраны; $g_L$ — проводимость утечки (*leak conductance*); $E_L$ — обратный потенциал утечки (*leak reversal potential*); $\Delta_T$ — коэффициент наклона (*slope factor*); $V_T$ — опорный потенциал (*reference potential*); $\tau_w$ — постоянная времени затухания (*reduction time constant*); $\alpha$ — переменная связи (*coupling variable*).

Для токов, входящих в сумму $I(t)$, формулируются отдельные уравнения.

Система дифференциальных уравнений (2.1), (2.2) обычно численно интегрируется простейшим методом Эйлера [7–9]. Путем замены производной на конечную разность из уравнения (2.1) для потенциала мембраны нейрона $V$ получается следующее рекуррентное выражение:

$$V_{n+1} = V_n + \frac{h}{C}\left[ g_L\Delta_T \exp\left(\frac{V_n - V_T}{\Delta_T}\right) - g_L\left(V_n - E_L\right) - \alpha\left(V_n - E_L\right) + I_n \right], \quad (2.3)$$

в котором ток адаптации $w \approx \alpha(V - E_L)$ взят из конечной разности уравнения (2.2) в подпороговой области ($V_n < V_T$) при малом изменении $w$ и большом временном шаге $h \gg \tau_w$

$$w_{n+1} = w_n + \frac{h}{\tau_w}\left[\alpha\left(V_n - E_L\right) - w_n\right] + b.$$

Изменение $w$ за шаг становится большим при генерации спайка ($b$ = 10–40 пА, $\alpha(V_n - E_L)$ = 0–40 мВ) [7]. Добавление $b$ происходит только при генерации спайка [10].

До превышения порога активации экспоненциальный член в (2.3) практически равен нулю во всех частях нейрона. При превышении порога активации генерация отдельного спайка в соме описывается экспоненциальным членом в выражении (2.3) [7]. В предлагаемой модели нейрона при превышении порога активации рассчитывается средняя частота

спайков, поэтому форма отдельных спайков не имеет значения. Важен сам факт генерации спайка в соме при превышении порога активации нейрона, т. е. экспоненциальный рост потенциала можно заменить на мгновенный рост, например, с помощью функции Хевисайда. Средняя частота потенциалов действия (число событий возникновения спайков за шаг моделирования) должна зависеть от постсинаптических потенциалов от дендритов.

На рис. 2.3, *а* показан результат расчета постсинаптического потенциала в дендрите по формуле (2.3), а на рис. 2.3, *б* — средней частоты потенциалов действия в соме в случае превышения порога генерации спайка. Пока постсинаптический потенциал остается выше порогового значения, сома продолжает генерировать импульсы, частота следования которых вычисляется как функция от величины превышения постсинаптического потенциала над порогом генерации спайка (25 мВ на рис. 2.3, *а*) и изменяется согласно графику на рис. 2.3, *б*.

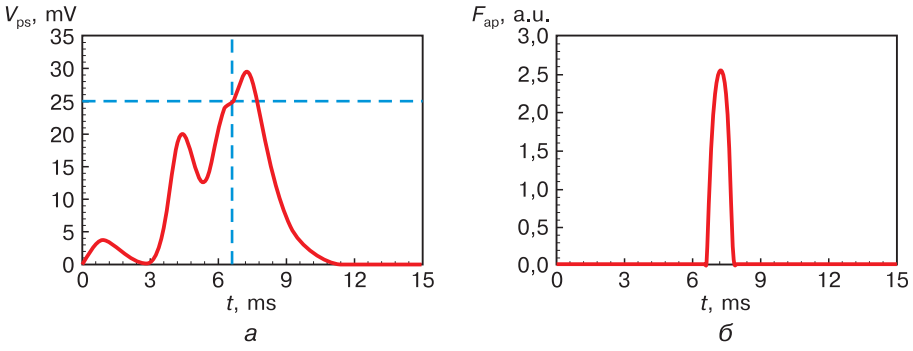



**Рис. 2.3.** Постсинаптический потенциал в дендрите (*а*), рассчитанный по (2.1), и соответствующая частота потенциалов действия в соме (*б*)

Учет экспоненциального члена в уравнении привел бы к генерации спайка при превышении порогового потенциала (в момент времени $t = 6{,}6$ мс на рис. 2.3, *а*). Так как генерация спайков происходит в аксонном холмике, который является частью сомы, исключение экспоненты из расчета потенциала дендрита оправдано.

Замена отдельных спайков на среднюю частоту дает возможность увеличить шаг по времени и, как следствие, увеличить скорость расчета нейросети. Чтобы дополнительно повысить эффективность вычислений сверхбольшой нейросети, имитирующей работу кортикальной колонки, на автономном аппаратном средстве с ограниченными вычислительными ресурсами переходим к описанию работы всех функциональных частей нейрона в виде унифицированного рекуррентного выражения. Для этого уравнение (2.3) представим в следующем виде:

$$V_{n+1} = aV_n - b + p_1 x^{p_2}, \qquad (2.4)$$

где суммарный ток через рецепторы $I = p_1 x_2^p$;

$$a = 1 - \frac{h}{C}(g_L + \alpha); \quad b = \frac{hE_L}{C}(g_L + \alpha).$$

Вид выражения для $I$ и константы $p_1$ и $p_2$, индивидуальные для каждой части нейрона и определяющие мгновенную реакцию на входной сигнал. Для дендрита они подобраны из аппроксимации экспериментальных данных измерения потенциала его мембраны [11]. На рис. 2.4 за возрастание рассчитанного по формуле (4) потенциала отвечает ток через рецепторы, а за релаксацию $V$ до нуля во времени — константа $b$.

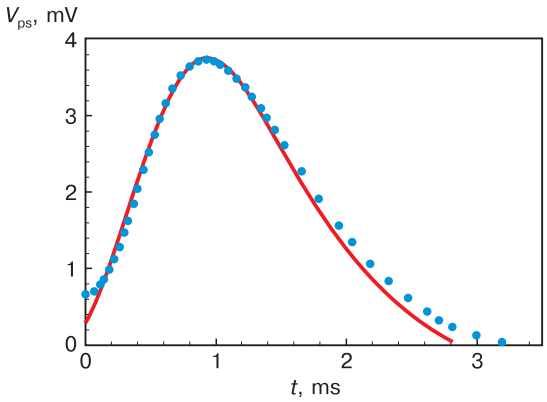



**Рис. 2.4.** Сравнение расчетного и экспериментально измеренного допорогового постсинаптического потенциала в дендрите на расстоянии 25 мкм от стимулирующего электрода [11]

Совпадение кривых достигается при следующих значениях коэффициентов в выражении (2.4): $p_1 = 0{,}305$; $p_2 = 0{,}5$; $a = 0{,}91$; $b = 0{,}09$. Максимальное отклонение расчетного потенциала от экспериментального составляет 0,37 мВ.

С помощью унифицированного рекуррентного выражения (2.4) кроме мембранного потенциала дендрита $V$, можно рассчитать количество рецепторов нейромедиатора на мембране дендрита, частоту следования потенциалов действия или фасилитацию аксона.

При численном расчете потенциалов с использованием конечной разности (2.3) временной шаг не может быть больше длительности потенциала действия в соме, который равен в среднем 2 мс. Использование большего временного шага приведет к искажению импульсов потенциала действия и увеличению их длительности. Переход к средней частоте потенциалов действия позволяет использовать больший размер шага, что приводит к сокращению времени расчета.

Ограниченные вычислительные возможности автономного средства вынуждают переходить к большему шагу, с учетом того, что общее время расчета сверхбольших нейросетей складывается из времени расчета каждого нейрона. При переходе к средней частоте потенциалов действия вместо учета отдельных спайков приходится расплачиваться частичной потерей информации о временных корреляциях активации нейронов, поскольку минимальная длительность между последовательными событиями равна длине временного шага. Однако если использовать большой шаг при спайковом кодировании, то это приведет к большей потере информации.

Алгоритм расчета прохождения сигнала через нейрон состоит из цикла расчета всех его дендритов, далее сомы, затем — цикла расчета его аксонов. Если нейрон активировался, то после расчета сомы в отдельном цикле пересчитывается число рецепторов на дендритах.

### 2.1.1. Модель дендрита

Основной частью является дендрит — именно его свойства меняются в ходе обучения. Кратковременная синаптическая пластичность описывается постепенным угасанием во времени вызванного постсинаптического потенциала в дендрите и фасилитации в аксоне. Появление потенциала в дендрите увеличивает на его синаптической мембране число переменных рецепторов нейромедиатора, вызвавшего его активацию. При отсутствии входных сигналов рецепторы со временем медленно распадаются. Так реализуется долговременная синаптическая пластичность — долговременное запоминание и забывание.

Число $M$ молекул нейромедиатора (возбуждающего $T_M = 1$ или угнетающего $T_M = -1$ типа), выброшенных пресинаптическим нейроном, сравнивается с числом соответствующих рецепторов ($R_+$ или $R\_$). Очевидно, числа $R_+$ и $R_$ограничивают максимальное число активированных рецепторов ($MR$). Новый вызванный постсинаптический потенциал дендрита является функцией числа активированных рецепторов и типа нейромедиатора. Число активированных рецепторов определяет синаптический ток заряда мембраны дендрита $p_1 MR_i^{p_2}$.

Полный постсинаптический потенциал дендрита $Vps_i$ на текущем шаге складывается от ПСП, полученного от рецепторов ($Vps_{\text{new}}$), и остаточного ($Vps_{\text{old}}$), который релаксирует к нулю по линейному закону со скоростью $Vps$, модулированному показательной функцией от времени. Учитывая, что постсинаптический потенциал $Vps$ может быть либо возбуждающим, либо тормозящим, в выражение (2.4) добавлен коэффициент $T_M \in \{-1, 1\}$, отвечающий за направление синаптического тока:

$$Vps_i = T_M\, p_1 MR_i^{p_2} + \left(Vps_{i-1} - Ups\right)a. \tag{2.5}$$

Блок-схема расчета постсинаптического потенциала в зависимости от числа выброшенных пресинаптическим нейроном молекул нейромедиатора по формуле (2.5) приведен на рис. 2.5.

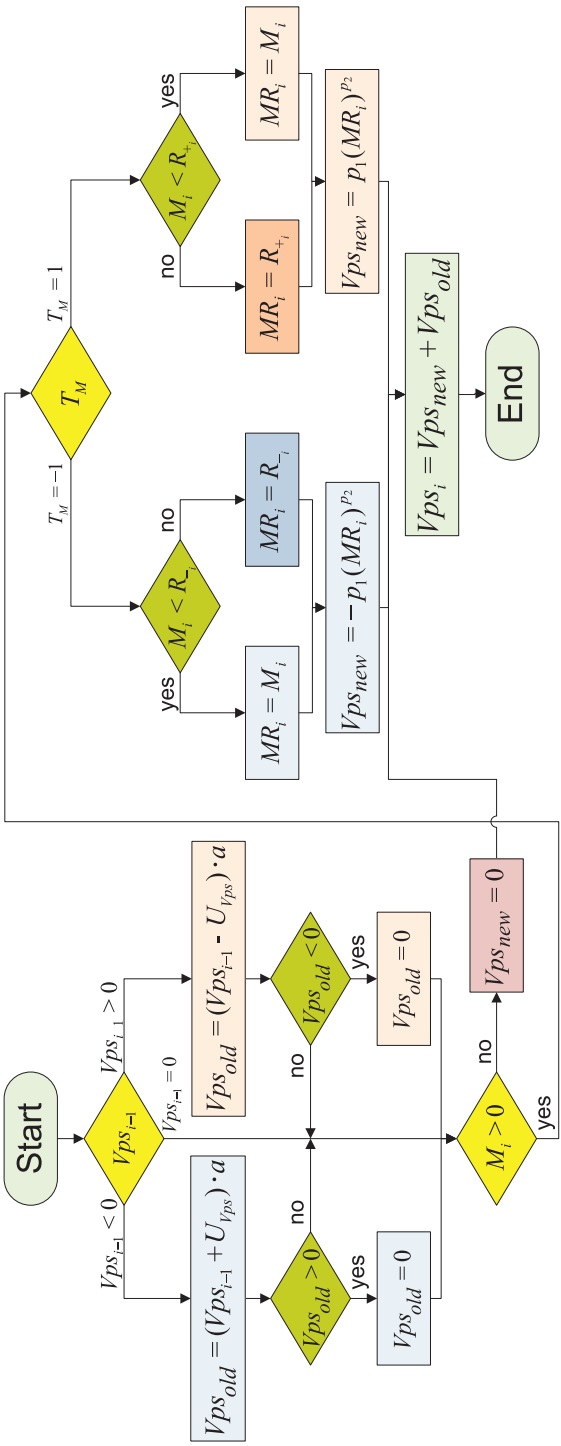



**Рис. 2.5.** Блок-схема расчета постсинаптического потенциала

Минимальные числа рецепторов обоих нейромедиаторов на мембране дендрита равны, соответственно, $R_{+\text{NMDA}}$ и $R_{-\text{GABAc}}$. Если максимальное число рецепторов $r_+$ ($r\_$) не достигнуто и при активации данного нейрона на данном дендрите присутствовал ПСП, то число новообразующихся рецепторов ($R_{+\text{AMPAnew}}$ и $R_{-\text{GABAnew}}$) вычисляется как функция от произведения частоты следования потенциалов действия $Fap_i$ и вызванного постсинаптического потенциала. Рецепторы разрушаются со временем по линейному закону (со скоростями $U_{R_+}$ и $U_{R\_}$), промодулированному показательной функцией. Общее количество переменных рецепторов нейромедиатора возбуждающего $R_{+\text{AMPA}_i}$ или тормозящего $R_{-\text{GABA}_{vi}}$ типов рассчитывается на основе выражения (2.4) следующим образом:

$$R_{+\text{AMPA}_i} = p_1\left(F_{\text{AP}_i} Vps_i\right)^{p_2} + \left(R_{+\text{AMPA}_{i-1}} - U_{R_+}\right)a; \tag{2.6}$$

$$R_{-\text{GABAv}_i} = p_1\left(F_{\text{AP}_i} Vps_i\right)^{p_2} + \left(R_{-\text{GABAv}_{i-1}} - U_{R\_}\right)a. \tag{2.7}$$

Итоговое число рецепторов равно сумме постоянного числа рецепторов ($R_{+\text{NMDA}}$ и $R_{-\text{GABA}_c}$), новообразованных и остаточного количества переменных рецепторов ($R_{+\text{AMPAold}}$ и $R_{-\text{GABA}_{v}\text{old}}$):

$$R_{+i} = R_{+\text{NMDA}} + R_{-\text{AMPA}_i}; \tag{2.8}$$

$$R_{-i} = R_{-\text{GABA}_c} + R_{-\text{GABA}_{vi}}. \tag{2.9}$$

Блок-схема алгоритма расчета количества рецепторов на мембране дендрита по формулам (2.6)–(2.9) представлена на рис. 2.6.

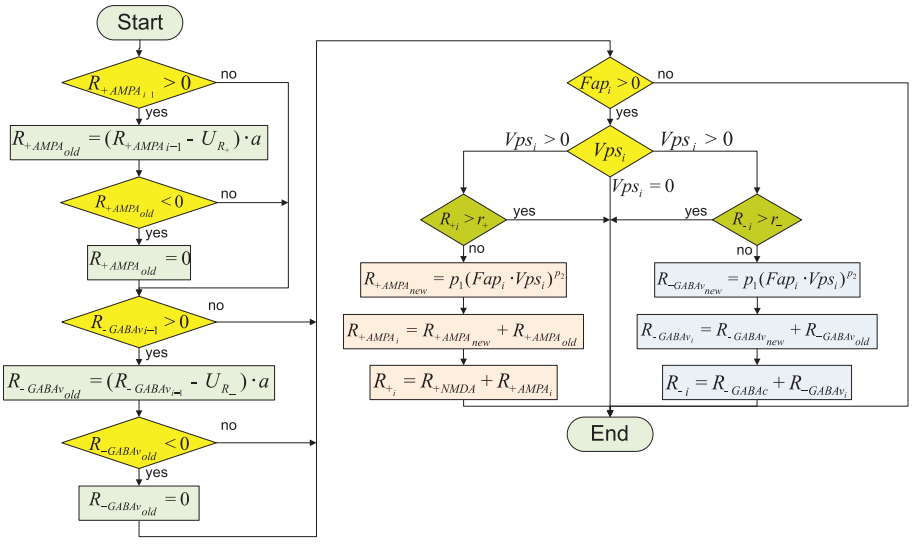



**Рис. 2.6.** Блок-схема расчета числа рецепторов на мембране дендрита

### 2.1.2. Модель сомы

В теле нейрона вызванные постсинаптические потенциалы всех дендритов складываются и сравниваются с пороговым значением $L_{\text{const}}$ (рис. 2.7). Когда суммарный потенциал от дендритов превышает порог срабатывания нейрона $L_{\text{const}}$, в аксонном холмике появляются импульсы потенциала действия, частота следования которых $F_{\text{AP}}$ является функцией от $U_p$ — суммы ПСП за вычетом порога:

$$F_{\text{AP}_i} = p_1 U_{p_i}^{p_2}. \qquad (2.10)$$

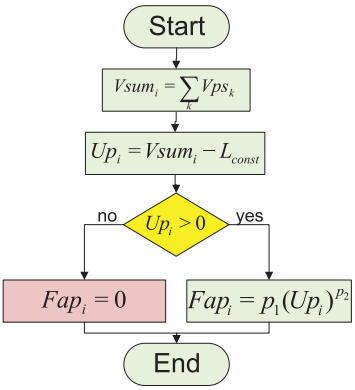



**Рис. 2.7.** Блок-схема расчета сомы

Выражение (2.10) получено из (2.4) при коэффициенте $a = 0$, поскольку генерация потенциалов действия происходит только при превышении суммы постсинаптических потенциалов $U_p$ порога активации $L_{\text{const}}$. Как только эта сумма становится ниже порога — генерация прекращается.

### 2.1.3. Модель аксона

Потенциалы действия из аксонного холмика вызывают в аксоне фасилитацию *FS*, величина которой пропорциональна концентрации ионов кальция (миллимоль/литр) в аксонной терминали и определяется как функция частоты потенциалов действия $F_{\text{AP}}$. Фасилитация определяет изменение реакции аксона в зависимости от прошедших через него импульсов: чем больше фасилитация, тем больше молекул нейромедиатора выбросит аксон в синаптическую щель при той же частоте потенциалов действия. Расчет фасилитации производится по формуле (2.4) при замене *V* на *FS*. На текущем временном шаге $FS_i$ полная фасилитация будет складываться из новой фасилитации $FS_{\text{new}}$ и оставшейся от предыдущих активаций $FS_{\text{old}}$. Остаточная фасилитация линейно релаксирует до нуля со скоростью $U_{\text{FS}}$:

$$FS_i = p_1 F_{\text{AP}_i}^{p_2} + \left( FS_{i-1} - U_{\text{FS}} \right) a. \qquad (2.11)$$

Расчет фасилитации по формуле (2.11) не приводит к большой ошибке. На рис. 2.8 представлено сравнение изменения концентрации ионов кальция в аксонной терминали после одиночного спайка (красная кривая) при расчете по формуле (2.11) и аналогичного изменения концентрации ионов, описанного в подробной модели [12].

Из рис. 2.8 следует, что уменьшение концентрации ионов кальция во времени практически совпадает для простой и сложной моделей. Причем такого совпадения достаточно, поскольку следующий импульс потенциала действия может появиться только в период рассасывания ионов кальция.

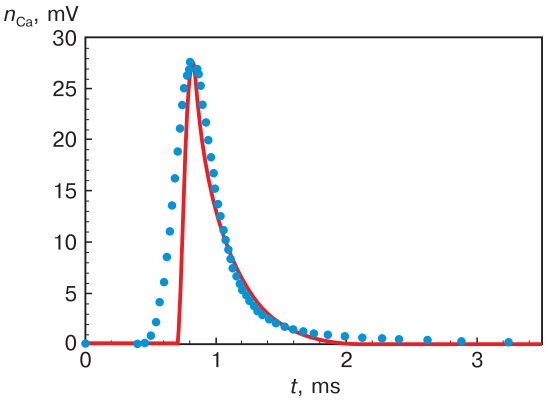



**Рис. 2.8.** Изменение концентрации ионов кальция в аксонной терминали после одиночного спайка:
*красная кривая* — расчет по формуле (2.11); *синяя кривая* — данные работы [12]

Итоговое количество молекул нейромедиатора $M$, выбрасываемого аксоном, является функцией суммы частоты следования потенциалов действия $F_{AP_i}$ и текущей фасилитации $FS_i$:

$$M_i = p_1(F_{\mathrm{AP}_i} + FS_i)_2^p. \qquad (2.12)$$

Выражение (2.12) рассчитывается так же, как (2.4), но с коэффициентом $a = 0$, поскольку длительность процесса выброса нейромедиаторов меньше, чем длина временного шага.

Блок-схема расчета прохождения сигнала через аксон по формулам (2.11) и (2.12) приведена на рис. 2.9.

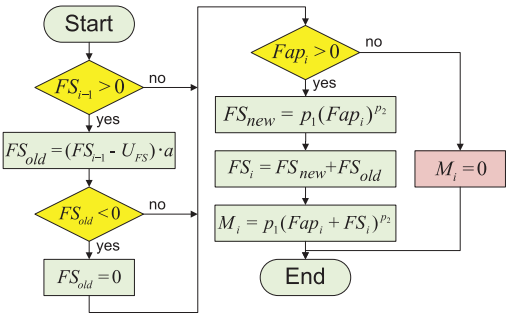



**Рис. 2.9.** Блок-схема расчета аксона

Следовательно, изменение сигнала во времени в модели нейрона описывается несколькими числами: числом молекул нейромедиаторов (в синаптической щели), вызванным постсинаптическим потенциалом (в дендрите), частотой потенциалов действия (в соме), формированием кальциевого

заряда в аксоне — фасилитацией. Свойства функциональных частей нейрона учитываются в формулах преобразования сигналов. Изменение динамических свойств дендрита и затухание сигнала реализовано в виде умножения линейной и показательной функций, монотонно убывающих во времени.

## 2.2. ПРИНЦИПЫ ПОСТРОЕНИЯ НЕЙРОСЕТИ НА ОСНОВЕ БИОМОРФНОЙ МОДЕЛИ НЕЙРОНА

Для построения тестовой нейросети были сформулированы два принципа. Первый принцип — это последовательная сборка нейросети из функциональных блоков с определенными функциями, включающих несколько связанных нейронов. Например, блоками задержки реализуется синхронизация сигналов от других блоков. Из таких блоков создаются более крупные структуры — кластеры, также выполняющие свои функции. Таким образом, нейронная сеть имеет модульную структуру.

Второй принцип — построение блоков основывается на обнаруженных реальных нейронных сетях, выполняющих определенные задачи в живом головном мозге. Связи между нейронами в сети задаются изначально на основе экспериментальных данных нейрофизиологии, а не формируются путем перестройки весовых коэффициентов равнозначных связей.

Покажем возможность создания нейросети по приведенным принципам на основе биоморфной модели нейрона. Функциональная схема тестовой нейросети, выполняющей простую ассоциацию вводимых символов во времени, представлена на рис. 2.10.

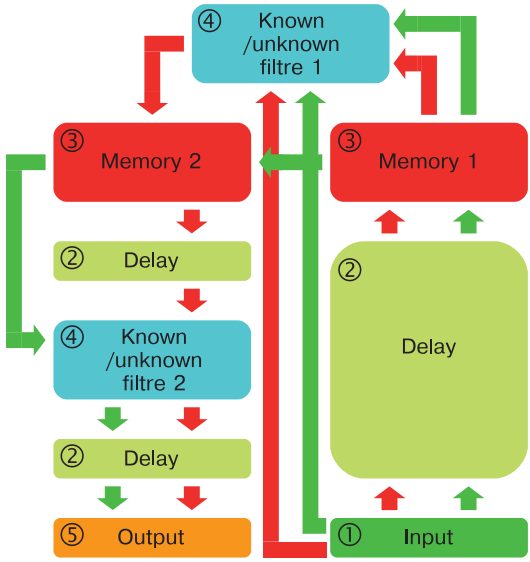



**Рис. 2.10.** Функциональная схема тестовой нейросети:
красный путь — при вводе неизвестного символа; зеленый — при вводе известного

Структурно архитектуру тестовой нейросети можно разделить на несколько блоков. Блоки ввода/вывода 1 и 5 преобразуют сигналы в требуемый формат, например, в символы. Программа расчета переводит вводимую в специальное поле символьную последовательность в последовательность активации соответствующих символам входных нейронов. Несколько блоков задержки 2 служат для синхронизации времени активации определенных нейронов, например, чтобы усилить связь между требуемыми нейронами в блоке запоминания 3 по правилу Хебба. В блоке запоминания последовательности пар введенных символов хранятся в виде усиленных связей. Передача входной последовательности символов на выход может происходить двумя путями: прямой передачей от входа и параллельным запоминанием пар, либо извлечением информации из параметров связей блока запоминания. Первый вариант выполняется при вводе неизвестной последовательности. Если вводится уже известная последовательность, то активируются нейроны из блока запоминания последовательности пар, ингибирующие прямую передачу посредством фильтра 4 и активирующие через усиленные связи нейроны, передающие затем на выход запомненные пары. Специальный фильтр 4 позволяет дополнить соответствующей парой только последний сигнал последовательности.

В режиме обучения сеть запоминает последовательность активации нейронов входного блока. Состав выходного блока 5 аналогичен входному 1. Активация входного нейрона поступившим сигналом в обученной нейросети вызывает активацию соответствующего ему нейрона и нейрона, активировавшегося вслед за ним на входе во время обучения.

Неизвестная последовательность активирующих сигналов проходит до конца неизменной, при этом в блоке запоминания запоминаются упорядоченные пары сигналов. Число рецепторов на мембране соответствующих дендритов нейронов из блока запоминания увеличивается таким образом, что повторный ввод первого символа пары вызовет передачу всей пары на выходной блок. Если же на вход подается новая последовательность сигналов, то уже вводившиеся последовательности дополняются в конце парным к последнему символу сигналом, неизвестные части запоминаются как обычно.

## 2.3. СИМУЛЯЦИЯ ТЕСТОВОЙ НЕЙРОСЕТИ

На рис. 2.11 представлен скриншот работы компьютерной программы, которая была специально разработана для моделирования тестовой нейросети.

Цвет нейрона соответствует типу сомы. Для разных типов сомы пороги суммы постсинаптических потенциалов, при превышении которого нейрон генерирует потенциал действия, приведены в табл. 2.1.

Всего в нейросети используется 11 типов связей (пар дендрит—аксон), характеристики которых приведены в табл. 2.2.

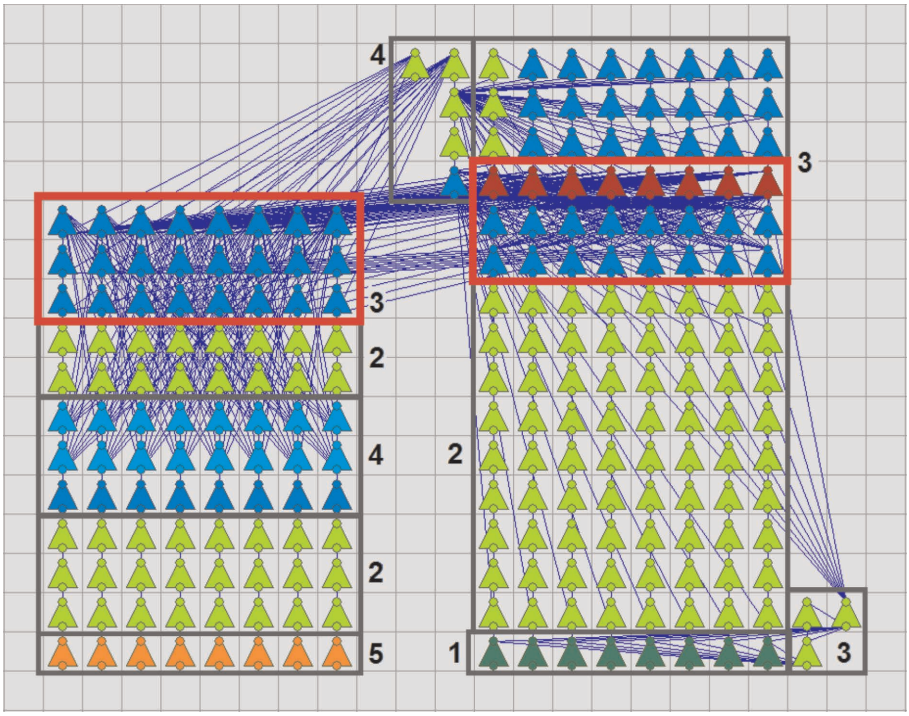



**Рис. 2.11.** Скриншот тестовой нейросети.
Нейроны обозначены треугольниками:
1 — входной блок; 2 — блоки задержки; 3 — блок запоминания;
4 — фильтр известного/неизвестного; 5 — выходной блок

*Таблица 2.1*

**Параметры разных типов сомы**

| Type | Input | Sum1 | Sum2 | Sum3 | Sum4 | Output |
|---|---|---|---|---|---|---|
| $L_{\text{const}}$ | — | 0 | 1 | 55 | 25 | 1 |
| $F_{\text{AP}}$ | 1 | 1 | 1 | $1,2\sqrt{U_{\text{pp}}}$ | $1,2\sqrt{U_{\text{pp}}}$ | 1 |

В процессе симуляции в качестве входных сигналов используются вводимые с клавиатуры символы. Каждый отдельный символ (цифры и символы «+», «=») соответствует своему нейрону входного блока. Таким образом, последовательность символов в строке интерфейсной программы преобразуется в последовательность активации нейронов входного блока. Каждый раз, когда выполняется тело основного цикла, программа искусственно активирует нейрон, соответствующий текущему символу во введенной строке. Выходной блок тестовой нейросети выполняет обратное

Таблица 2.2

**Параметры синаптических связей**

| Part | Parameter | | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ | $S_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Axon | $T_M$ | | + | – | + | + | + | – | + | + | – | + | + |
| | $FS_{\text{new}}$ | $p_1$ | 1 | 100 | 1 | 2,7 | 1 | 1 | 5 | 1 | 100 | 2,7 | 5 |
| | | $p_2$ | 1 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 0,7 |
| | | $U_{\text{FS}}$ | 0 | 0 | 0 | 100 | 0 | 0 | 10 | 0 | 0 | 100 | 10 |
| | | $a_{\text{FS}}$ | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 2 |
| | $M$ | $p_1$ | 1 | 100 | 1 | 2,7 | 1 | 1 | 5 | 1 | 100 | 2,7 | 5 |
| | | $p_2$ | 1 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 0,7 |
| Dendrite | $PSP_{\text{new}}$ | $p_1$ | 1 | 4 | 2 | 5 | 1,5 | 10 000 | 14,5 | 30 | 10 000 | 9 | 4,8 |
| | | $p_2$ | 1 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 1 | 1 | 0,7 | 0,7 |
| | | $U_{\text{PSP}}$ | 1 | 10 | 25 | 150 | 1000 | 20 000 | 100 | 100 | 20 000 | 150 | 100 |
| | | $a_{\text{PSP}}$ | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| | $R_{+AM}$ | $p_1$ | 1 | 0 | 0 | 3,9 | 2 | 0 | 1,2 | 0 | 0 | 4,9 | 0 |
| | | $p_2$ | 0 | 0 | 0 | 0,9 | 0 | 0 | 0,5 | 0 | 0 | 0,9 | 0 |
| | | $U_{R_+}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | $a_{R_+}$ | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| | $R_{-\text{GABA}v\text{new}}$ | $p_1$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | $p_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | $U_{R_-}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | $a_{R_-}$ | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| | $R_{+\text{NMDA}}$ | | 1 | 1 | 1 | 0,1 | 1 | 0 | 20 | 1 | 0 | 0,1 | 10 |
| | $R_{-\text{GABAc}}$ | | 1 | 1 | 1 | 0,1 | 1 | 1 | 0,1 | 0 | 1 | 0,1 | 0,1 |

преобразование — активация нейрона вызывает появление в строке вывода интерфейсной программы соответствующего ему символа. Ниже приведена последовательность вводимых строк и ответов, полученных от тестовой нейросети:

$$1+2=3\rightarrow 1+2=3;$$

$$2+1=3\rightarrow =32+1=;$$

$$2+1=3\rightarrow 2+1=3;$$

$$1+1+1=\rightarrow 1+1+1=3.$$

Первые три строки являются обучающими последовательностями. Последний ввод и ответ на него показывает реакцию обученной нейросети на неизвестную последовательность в виде корректного дополнения строки символом. Нейросеть не выполняет арифметические операции, а только дополняет строку символом, который чаще других встречался после символа «=» при ее работе.

Приведенный простой пример свидетельствует о работоспособности тестовой нейросети, построенной с использованием биоморфной модели нейрона и выполняющей простую ассоциацию вводимых символов во времени.

## 2.4. АДАПТАЦИЯ БИОМОРФНОЙ НЕЙРОСЕТИ К АППАРАТНОЙ ЧАСТИ НЕЙРОПРОЦЕССОРА

Адаптация к разработанному нейропроцессору биоморфной нейросети, построенной по изложенным принципам, по существу сводится к адаптации программной биоморфной модели нейрона, поскольку информация о связях между нейронами переносится непосредственно. Такая адаптация заключается в пересчете коэффициентов в формулах биоморфной модели нейрона через электрические параметры узлов нейропроцессора. Для этого необходимо произвести следующие процедуры:

1) соотнести состояния комплементарных мемристоров с числом рецепторов нейромедиаторов;
2) установить связь скорости затухания остаточного постсинаптического потенциала в дендрите со скоростью разряда суммирующего конденсатора в электрической схеме нейрона (см. рис. 1.18);
3) пересчитать коэффициенты скоростей синтеза и распада рецепторов в синапсе;
4) подобрать величину порогового напряжения генерации импульса в аппаратном нейроне, соответствующую порогу активации нейрона в биоморфной модели.

Первая процедура предполагает пересчет числа активированных рецепторов $MR_i$ через состояние мемристоров аппаратного синапса в формуле

расчета постсинаптического потенциала нейросети (2.5). Пересчитываемая величина активированных рецепторов $MR_i$ зависит от сопротивлений мемристоров в открытом и закрытом состояниях и от коэффициента усиления источника тока, управляемого током (ИТУТ). Значения коэффициентов $p_1$ и $p_2$ задаются схемой преобразования тока в ИТУТ.

Во второй процедуре настраивается скорость разряда суммирующего конденсатора в аппаратном нейроне через скорость затухания постсинаптического потенциала $U_{\text{ps}}$ в формуле (2.5) путем выбора сопротивления резистора *RC*-цепи. Скорость разряда выбирается такой, чтобы обеспечить время полного разряда конденсатора равным времени полного исчезновения постсинаптического потенциала в биоморфной модели нейрона при отсутствии входных импульсов.

Третья процедура заключается в пересчете скоростей распада возбуждающих и тормозящих нейромедиаторов $U_{R+}$, $U_{R\_}$ через частоту генератора релаксации (безусловного разобучения) аппаратных синапсов. Эти скорости распада содержатся в формулах (2.6) и (2.7), описывающих изменение числа рецепторов нейромедиаторов.

Порог активации нейрона в последней процедуре однозначно связан с порогом напряжения на конденсаторе, при котором происходит генерация выходного импульса аппаратного нейрона. Так как суммирование постсинаптических потенциалов в биоморфной нейросети выполняется по формуле

$$F_{\text{AP}_i} = p_1 \left( \sum_k Vps_k \right)_i^{p_2}, \tag{2.13}$$

то порог активации аппаратного нейрона $V_{\text{hw}}$ может быть вычислен исходя из порога активации нейрона в исходной нейросети $V_{\text{sw}}$ и значений параметров $p_1$ и $p_2$

$$V_{\text{hw}} = p_1 V_{\text{sw}}^{p_2}. \tag{2.14}$$

Таким образом, все основные функции биоморфной модели нейрона (2.5)–(2.7), (2.13) и (2.14) присутствуют в аппаратной реализации нейропроцессора.

## 2.5. ПРОГРАММНО-АППАРАТНАЯ РЕАЛИЗАЦИЯ НЕЙРОСЕТИ

Основная идея заключается в том, чтобы в алгоритме работы тестовой нейросети между программными нейронами реализовать связи, усиливающиеся во время активации, путем применения аппаратного блока с мемристорами. Предложенная программная биоморфная модель нейрона позволяет без существенного изменения алгоритма заменить в нем часть расчета количества рецепторов мембраны дендрита на эффект увеличения электрической

проводимости мемристора при подаче на него импульсов электрического напряжения. В этом случае в мемристоре можно получить некоторое количество промежуточных резистивных состояний. Благодаря этому имеется возможность реализации механизма синаптической пластичности, который заложен в модель тестовой нейросети в блоке запоминания, показанный на рис. 2.11.

Множество резистивных состояний в мемристоре можно получить разными способами, в том числе и при использовании импульсного программирования. Для этого требуется подавать на мемристор короткие по времени (длительностью во много раз меньше времени полного перепрограммирования мемристора) импульсы заданной длительности с амплитудой напряжения выше порога его программирования. Увеличение удельной электропроводности мемристора $\Delta\sigma$ за один короткий импульс, поданный в интервале времени от $t_1$ до $t_2$, будет прямо пропорционально прошедшему через него электрическому заряду, что описывается классической формулой $\Delta\sigma = kq$. Учитывая, что электрический ток, протекающий через мемристор, определяется выражением $I(t) = \mathrm{d}q/\mathrm{d}t$, то изменение электропроводности мемристора можно задать формулой:

$$\Delta\sigma = k\int_{t_1}^{t_2} I(t)\,\mathrm{d}t, \tag{2.15}$$

где $k$ — коэффициент пропорциональности, зависящий от свойств мемристорного материала (в общем случае $k$ является зависимой величиной от текущей проводимости мемристора $y$, но в линейном упрощении при малых изменениях электропроводности, $k$ можно считать константой); $I(t)$ — функция, описывающая импульс тока программирования.

Предполагая дискретность изменения состояний мемристора от $i - 1$ импульса к $i$ импульсу, его текущую электропроводность $y_i$ можно выразить через предыдущую $\sigma_{i-1}$ формулой $\sigma_i = \sigma_{i-1} + \Delta\sigma$. Интеграл в формуле (2.13) можно упростить, допустив, что ток во время короткого импульса программирования является постоянным и равным $I(t) = U_\text{р}\sigma_{i-1}$, где $U_\text{р}$ — напряжение программирования. Таким образом, при подаче короткого импульса программирования с длительностью $\tau$, формулу (2.13) можно свести к рекуррентному выражению:

$$\sigma_i = \sigma_{i-1} + kU_\text{р}\tau\sigma_{i-1}, \tag{2.16}$$

где $\sigma_i$ — удельная электропроводность после подачи $i$ программирующих импульсов; $\sigma_{i-1}$ — удельная электропроводность после подачи $i - 1$ программирующих импульсов; $U_\text{р}$ — напряжение программирования.

Последняя зависимость показывает, что электропроводность мемристора увеличивается при подаче некоторого количества коротких импульсов, что можно использовать в качестве увеличения количества рецепторов нейромедиатора. Следовательно, выражение (2.6) для общего количества

переменных рецепторов нейромедиатора возбуждающего $R_{+\mathrm{AMPA}_i}$ типа соответствует выражению (2.14) при следующих заменах

$$R_{+\mathrm{AMPA}_i} \to \sigma_i;$$

$$p_1\left(F_{\mathrm{AP}_i} Vps_i\right)^{p_2} \to kU_{\mathrm{p}}\tau\sigma_{i-1};$$

$$R_{+\mathrm{AMPA}_{i-1}} a \to \sigma_{i-1}.$$

Таким образом, процесс изменения проводимости мемристора моделирует процесс усиления нейронной связи за счет увеличения количества рецепторов мембраны дендрита при ширине импульса, пропорциональной $(F_{\mathrm{AP}_i} Vps_i)^{p_2}$.

Для исследования возможности замены программных алгоритмов расчета синапсов нейросети на мемристоры был создан испытательный стенд (см. рис. 1.10), представляющий собой мемристорную микросхему, соединенную с классическим микроконтроллером КМОП-типа [13].

На схеме испытательного стенда (рис. 2.12) показано соединение портов ввода—вывода микроконтроллера с мемристорным кроссбаром размерностью 16 × 16 и подключение устройства к компьютеру через интерфейс USB. В микроконтроллер записана микропрограмма, реализующая пользовательский интерфейс, программную часть тестовой нейросети и алгоритм вывода—ввода импульсных сигналов для кроссбара.

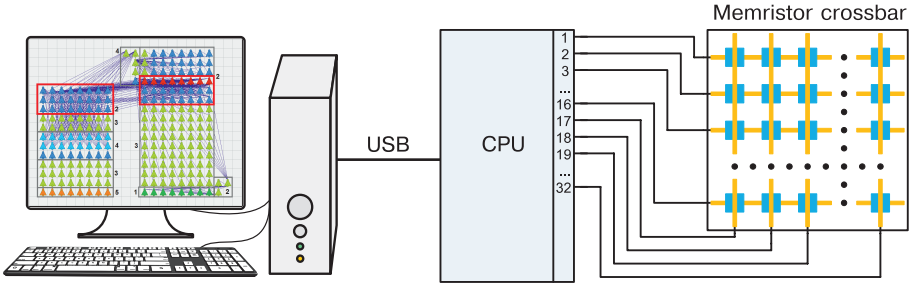



**Рис. 2.12.** Схема испытательного стенда

Импульсы для записи и чтения мемристоров разной длительности и амплитуды формируются за счет особенностей электрической схемы порта микроконтроллера, позволяющей на выходе получить: высокий уровень напряжения (H-состояние, соответствующее 3 вольтам), низкий уровень напряжения (L-состояние, соответствующее 0 вольт) и высокоимпедансное Z-состояние, соответствующее отключению порта от мемристора. В схеме каждого порта микроконтроллера находятся четыре комплементарных полевых транзистора для вывода и один транзистор ввода логических уровней напряжения. Схема соединений транзистора показана на рис. 2.13. Программно управляемые состояния на линиях IN, OUT и ENABLE задают

три состояния порта: H, L и Z. Комбинацией состояний двух портов можно обеспечить импульсы как в одну, так и в другую сторону напряжения обеих полярностей, что необходимо для изменения состояния мемристоров.

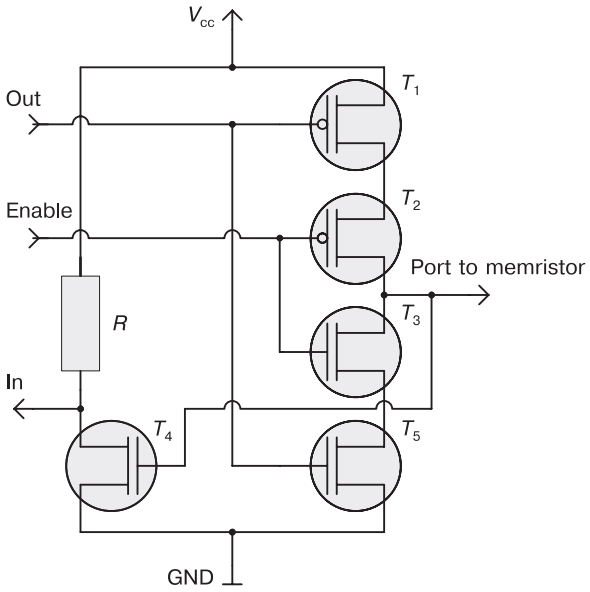



**Рис. 2.13.** Схема порта микроконтроллера:
$T_1$–$T_2$ — NMOS FETs; $T_3$–$T_5$ — PMOS FETs; $R$ — подтягивающий резистор

Каждый мемристор кроссбара подключен к двум портам микроконтроллера, на которых изначально установлено состояние L, что соответствует отсутствую напряжения на мемристоре. Запись и перезапись мемристоров осуществляется импульсами, формируемыми за счет изменения состояния с L на H на одном из двух портов. Импульсы, подаваемые таким образом на мемристоры, имеют амплитуду напряжения ($U_p$) равную значению напряжения питания микроконтроллера. Это напряжение является настраиваемым параметром в пределах 2–5 В, путем изменения напряжения питания микроконтроллера.

Считывание текущего состояния мемристора выполняется посредством короткого импульса напряжения, сгенерированного на одном из подключенных к нему портов микроконтроллера. При этом другой подключенный к мемристору порт устанавливается в Z-состояние и настраивается на ввод логических сигналов. Проходящие через мемристоры импульсы заряжают внутреннюю емкость порта микроконтроллера через электрическое сопротивление мемристора. При достижении напряжения на этой емкости порога срабатывания микроконтроллер зафиксирует появление на порту логической единицы. Проводимость мемристора определяется по времени появления логической единицы, после подачи фронта (*rising*) считывающего импульса.

Для исключения взаимовлияния мемристоров, соединенных по схеме кроссбара, в эксперименте были использованы только несвязанные мемристоры в качестве части синаптических связей блока запоминания тестовой неросети, представленной на рис. 2.11.

Результаты прохождения сигналов из контроллера передавались в специально созданную компьютерную программу и показывались на мониторе в графическом виде прохождения импульсов по тестовой нейронной сети.

Для задействования мемристоров в нейросети необходимо подобрать длительность и амплитуду импульса, обеспечивающего гарантированное переключение состояния. Амплитуда импульса должна превышать пороговые значения напряжений перехода в низкопроводящее и высокопроводящее состояния мемристоров. Из характеристик мемристоров [14] в кроссбаре испытуемого аппаратного средства следует, что амплитуда переключающего импульса должна превышать 2,5 В. Для выбора оптимальных параметров работы мемристоров проведено исследование влияния длительности и амплитуды импульса на их удельное сопротивление.

В работе [15] исследована зависимость вольт-амперной характеристики мемристоров на диоксиде титана от длительности единичного импульса прямоугольной формы в пределах от 110 мкс до 33 с. Однако в живой нейронной сети измеренная средняя длительность импульсов составляет около 1–4 мс [16]. На рис. 2.14 представлены зависимости измеренных параметров мемристоров (вольтамперной характеристики и удельного сопротивления ρ) от длины импульса, охватывающей данный временной интервал. Вольтамперная характеристика измерялась при подпороговых напряжениях после воздействия единичного надпорогового импульса разной длительности и амплитудой 3 В. Относительная ошибка в измерении составила не более 4 %.

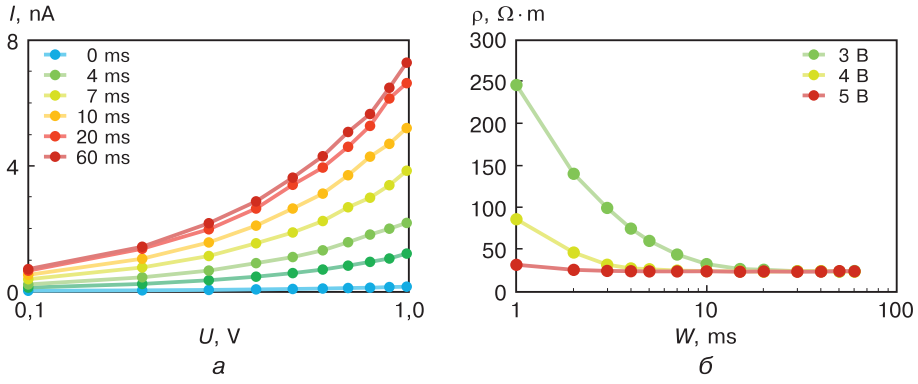



**Рис. 2.14.** Зависимости измеренных параметров мемристоров:

*а* — вольт-амперные характеристики мемристоров после прохождения надпорогового импульса 3 В; *б* — изменение удельного сопротивления мемристоров в зависимости от длительности и амплитуды надпорогового импульса

Из рис. 2.14, *а* следует, что при длительности 3-вольтового импульса 20 мс происходит практически полное переключение мемристора в низкоомное состояние. Полное переключение происходит за 60 мс, что соответствует времени переключения из низкопроводящего в высокопроводящее состояние на левой ветви ВАХ, приведенной в [14]. На рис. 2.14, *б* проиллюстрировано насыщение удельного сопротивления при амплитуде импульса 3—5 вольт в диапазоне длительностей импульсов реальных нейронов (1–4 мс). Учитывая, что в нейросети синапс заменен на единичный мемристор, нескольким синаптическим состояниям должны соответствовать несколько состояний мемристора. Полученная зависимость проводимости мемристора от числа поданных 1 мс импульсов (рис. 2.15) имеет линейный участок, который дает возможность использовать ряд равноотстоящих устойчивых состояний. При уменьшении амплитуды надпорогового импульса насыщение наступает за большее число импульсов, что позволяет реализовать больше синаптических состояний.

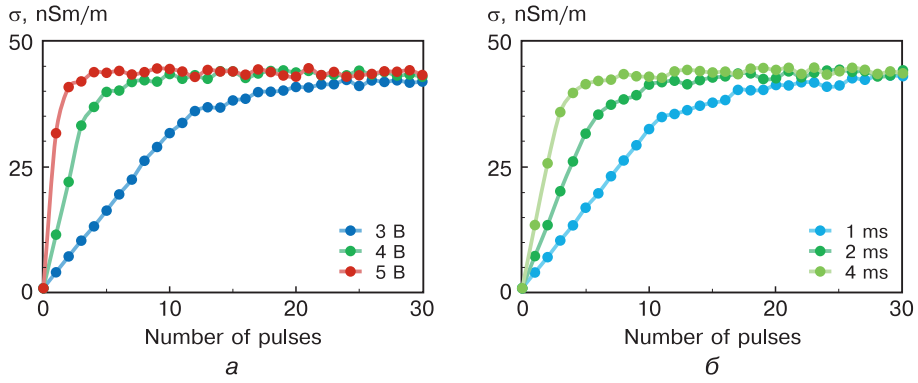



**Рис. 2.15.** Зависимость проводимости мемристора:
*а* — от числа поданных 1 мс импульсов при разных напряжениях;
*б* — от числа поданных 3 В импульсов разной длительности

Вид зависимости на рис. 2.15 практически совпадает с гиперболическим тангенсом, который часто используют в качестве активационной функции в нейросетях второго поколения. Таким образом, использование мемристоров позволяет автоматически учесть уменьшение чувствительности нейронов к входным сигналам, не производя вычисления в микроконтроллере.

Результаты, представленные на рис. 2.14 и рис. 2.15, показывают наличие в мемристоре с большим временем полного переключения (60 мс) множества промежуточных состояний с разной проводимостью между предельными высокопроводящим и низкопроводящим состояниями. Эти состояния можно использовать в процессах ассоциативного обучения нейросети на основе мемристорных синапсов и одновременной обработки входных

импульсов, заключающейся в их взвешивании и последующим суммированием в нейроне. Такой мемристор может взвешивать напряжения сигналов и эффективно обучаться при длительности импульсов, сопоставимой с длительностью потенциала действия реальных нейронов (1–4 мс).

## 2.6. ОСОБЕННОСТИ РАБОТЫ НЕЙРОСЕТИ НА ЭЛЕКТРОННОМ УСТРОЙСТВЕ С ЭНЕРГОНЕЗАВИСИМОЙ ПАМЯТЬЮ

Для моделирования работы биоморфной нейросети изготовлено электронное устройство [13], сочетающее в себе программируемые микроконтроллеры и энергонезависимую мемристорную память, и имеющее возможность работы совместно с персональным компьютером. В качестве первоначальной архитектуры выбраны: однослойный персептрон для первичного ассоциирования входных данных и биоморфная нейросеть.

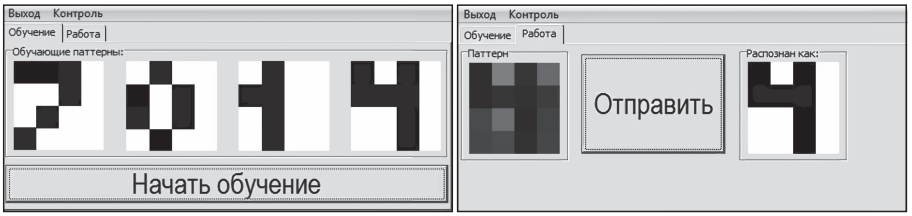
Встроенный персептрон — нейросеть для распознавания изображений, выполняющая первичную ассоциацию, качественно отличается от нейросети с алгоритмом линейной классификации, использованной в [17]. Ключевыми отличиями являются: большее число мемристоров и пикселей для распознавания образов; возможность распознавания полутоновых изображений; автономность и мобильность нейроморфного устройства в целом; автоматизированное обучение через программный интерфейс. Дальнейшее ассоциирование выполняет уже биоморфная нейросеть. Подобное сращивание различных архитектур ИНС используется для увеличения точности распознавания и ускорения вычислений [18–22].

Для тестирования персептрона была выбрана задача распознавания и классификации изображений [23]. На вход подавались монохромные полутоновые изображения цифр размером 4 × 4 пикселей. Процедура обучения заключается в подстройке весовых коэффициентов синаптических связей по специальным правилам, до тех пор, пока при предъявлении идеального изображения на выходе нейросети, соответствующим этому изображению не будет появляться максимальный по величине сигнал, а на остальных — минимальный. Обученная нейросеть продемонстрировала уверенное распознавание сильно зашумленных символов.

На рис. 2.16, *а* представлены шаблоны изображений нескольких символов, на которых обучалось электронное устройство, а на рис. 2.16, *б* продемонстрирован распознанный зашумленный полутонами символ.

Нейросеть реализована на языке C++, с использованием библиотек параллельного программирования для многоядерных систем с общей памятью. Адаптированная к электронному устройству версия биоморфной программы написана на языке assembler по принципу конечного автомата, что является перспективным подходом для алгоритмической реализации искусственного интеллекта. Причин перехода от C++ к ассемблеру

несколько: объектно-ориентированный подход, использованный в исходной программе, замедляет вычисления; такой подход неприменим к программированию используемого микроконтроллера. Кроме этого, написание кода на assembler позволяет достичь максимально возможной производительности. Программа верхнего уровня для ПК, управляющая электронным устройством с помощью виртуального USB-COM-порта, была написана на языке Delphi. Для этого языка существует библиотека, реализующая необходимые функции ввода и вывода информации.




*а* *б*

**Рис. 2.16.** Обучение нейросети (*а*)
и распознанный зашумленный полутонами символ (*б*)

При адаптации нейросетевых программ в первую очередь понадобилось написать процедуры записи и считывания состояния мемристоров. Проводники кроссбара соединены с портами микроконтроллеров, отвечающих за вход и выход. Считывание текущего состояния происходит посредством импульса напряжения с амплитудой ниже порога переключения мемристора. Запись осуществляется подобными импульсами с напряжением, превышающим порог.

Процедуры обучения персептрона и биоморфной нейросети различны. Персептрон обучается с учителем, т. е. веса связей подстраиваются на основании разницы между текущими значениями на выходе с желаемыми. Если на выходе значение выше требуемого, то его входные связи ослабляются пропорционально величине несоответствия. В случае недостаточного сигнала на выходе — соответствующие веса увеличиваются. Персептрон обучался посредством специальной программы с интерфейсной программой верхнего уровня, выполняемой на ПК.

Биоморфная сеть обучается согласно правилу Хебба: связь между одновременно активировавшимися нейронами усиливается. Ассоциативное основание, моделируемое архитектурой биоморфной нейросети, в процессе обучения запоминает последовательность сигналов в виде упорядоченных пар. В дальнейшем, при поступлении какого-либо сигнала, он проходит до выхода и дополняется сопряженным. Это один их видов ассоциаций, реализуемых в нервных узлах животных.

Таким образом, сначала мы обучаем персептрон, а потом его выход соединяем со входом биоморфной нейросети (см. рис. 2.11). Чтобы работа комбинированной нейросети была корректной, необходимо преобразовать

выходные сигналы персептрона к виду, требуемому для входных нейронов биоморфной нейросети. Каждый входной нейрон ассоциативного основания отвечает за определенный символ, и описывается возможными значениями {0; 1}. Поэтому для преобразования была использована функция Хевисайда.

При детальном анализе работы ассоциативного основания было выяснено, что синаптические связи, не участвующие непосредственно в запоминании и воспроизведении информации, можно описать функцией Хевисайда. Часть оставшихся синапсов были заменены на мемристоры. Эти два обстоятельства значительно сократили вычисления, осуществляемые для каждого временного шага, что увеличило быстродействие устройства в целом.

Список литературы


1. *Liu Ju.*, *Huang Y.*, *Luo Yu.* et al. Bio-inspired fault detection circuits based on synapse and spiking neuron models // Neurocomputing. 2019. Vol. 331. Pp. 473–482.
2. *Liu Ju.*, *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. Vol. 29. No. 4. Pp. 1287–1300.
3. *Liu Ju.*, *Mcdaid L.J.*, *Harkin J.* et al. Exploring self-repair in a coupled spiking astrocyte neural network // IEEE Transactions on Neural Networks and Learning Systems. 2018. Vol. 30. No. 3. Pp. 865–875.
4. *Филиппов В.А.* Моделирование нейрона // Александров Ю.И., Анохин К.В., Безденежных Б.Н. и др. Нейрон. Обработка сигналов. Пластичность. Моделирование. Фундаментальное руководство. Разд. 5 / Под ред. Е.Н. Соколовой, В.А. Филиппова, А.М. Черноризова. — Тюмень: Изд-во Тюменского гос. ун-та, 2008. — С. 468–535.
5. *Filippov V.A.*, *Bobylev A.N.*, *Busygin A.N.* et al. A biomorphic neuron model and principles of designing a neural network with memristor synapses for a biomorphic neuroprocessor // Neural Computing and Applications. 2020. Vol. 32. Pp. 2471–2485.
6. *Brette R.* Philosophy of the spike: Rate-based vs. spike-based theories of the brain // Frontiers in Systems Neuroscience. 2015. Vol. 9. Pp. 151.
7. *Malebra P.*, *Rulkov N.F.*, *Bazhenov M.* Large time step discrete-time modeling of sharp wave activity in hippocampal area CA3 // Commun Nonlinear Sci. Numer Simulat 2019. Vol. 72. Pp. 162–175.
8. *Rulkov N.F.*, *Neiman A.B.* Control of sampling rate in map-based models of spiking neurons // Commun Nonlinear Sci Numer Simulat 2018. Vol. 61. Pp. 127–137.
9. *Komarov M.*, *Krishnan G.*, *Chauvette S.* et al. New class of reduced computationally efficient neuronal models for large-scale simulations of brain dynamics // Journal of Computational Neuroscience. 2018. Vol. 44. Pp. 1–24.
10. *Touboul J.*, *Brette R.* Dynamics and bifurcations of the adaptive exponential integrate-and-fire model // Biological Cybernetics. 2008. Vol. 99. Pp. 319–34.
11. *Winters B.D.*, *Jin S.-X.*, *Ledford K.R.*, *Golding N.L.* Amplitude normalization of dendritic EPSPs at the soma of binaural coincidence detector neurons of the medial superior olive // Journal of Neuroscience. 2017. Vol. 37. No. 12. Pp. 3138–3149.
12. *Timofeeva Yu.*, *Volynski K.E.* Calmodulin as a major calcium buffer shaping vesicular release and short-term synaptic plasticity: facilitation through buffer dislocation // Frontiers in Cellular Neuroscience. 2015. Vol. 9. Pp. 239.

13. *Bobylev A.N.*, *Busygin A.N.*, *Pisarev A.D.* et al. Neuromorphic coprocessor prototype based on mixed metal oxide memristors. // International journal of nanotechnology. 2017. Vol. 14. No. 7/8. Pp. 698–704.
14. *Bobylev A.N.*, *Udovichenko S.Yu.* The electrical properties of memristor devices TiN/$Ti_xAl_{1-x}O_y$/TiN produced by magnetron sputtering // Russian Microelectronics. 2016. Vol. 45 No. 6. Pp. 396–401.
15. *Pickett M.*, *Strukov D.*, *Borghetti J.* et al. Switching dynamics in titanium dioxide memristive devices // Journal of Applied Physics. 2009. Vol. 106. No. 7. P. 074508.
16. *Bean B.* The action potential in mammalian central neurons // Nature Reviews Neuroscience. 2007. Vol. 8. Pp. 451–465.
17. *Prezioso M.*, *Merrikh-Bayat F.*, *Hoskins B.D.* et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors // Nature. 2015. Vol. 521. Pp. 61–64.
18. *Chelsia A.D.*, *Dargham J.A.*, *Chekima A.*, *Omatu S.* Combining neural networks for skin detection. Signal and image processing // Signal and image processing: An International Journal. 2010. Vol. 1. No. 2. Pp. 1–11. arXiv: 1101.0384.
19. *Plahl C.*, *Kozielski M.*, *Schluter R.*, *Ney H.* Feature combination and stacking of recurrent and non-recurent neural networks for LVCSR // IEEE ICASSP. 2013. Pp. 6714–6718.
20. *Wen C.*, *Rebelo A.*, *Zhang J.*, *Cardoso J.* A new optical music recognition system based on combined neural network // Pattern Recognition Letters. 2015. Vol. 58. Pp. 1–7.
21. *Ding Y.R.*, *Cai Y.J.*, *Sun P.D.*, *Chen B.* The use of combined neural networks and genetic algorithms for prediction of river water quality // Journal of Applied Research and Technology. 2014. Vol. 12. Pp. 493–499.
22. *Güler I.*, *Übeyli E.D.* ECG beat classifier designed by combined neural network model // Pattern Recognition. 2005. Vol. 38. Pp. 199–208.
23. *Бусыгин А.Н.*, *Писарев А.Д.*, *Кузьменко А.Ю.*, *Филиппов В.А.* Особенности моделирования работы биоморфной нейросети на электронном устройстве с энергонезависимой памятью и низким потреблением энергии // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2016. № 1. С. 92–100.

# ГЛАВА 3

# ЗАПОМИНАЮЩАЯ МАТРИЦА НА ОСНОВЕ КОМБИНИРОВАННОГО МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА

Существующие энергонезависимые запоминающие устройства на основе ячеек биполярного [1; 2] и униполярного [3] резистивного переключения используются только для хранения двоичных данных, и архитектура устройства не позволяет выполнять какую-либо обработку входных сигналов.

В настоящем параграфе описывается биоморфный подход к построению сверхбольшой запоминающей матрицы нейропроцессора, принципы которого не требуют строгой идентичности запоминающих элементов и высокой скорости переключения мемристорных ячеек. Ячейки запоминающей матрицы обеспечивают три операции над входными сигналами: сложение (+1), вычитание (–1) и непрохождение (0), причем использование промежуточных состояний мемристора позволяет выполнять взвешенное суммирование. Такое аналоговое суммирование требует намного меньшего числа элементов по сравнению с двоичными вычислениями.

Неидентичность элементов цифровой запоминающей матрицы приводит к ее неработоспособности, поскольку считывание происходит последовательно по адресу. Предлагаемая биоморфная запоминающая матрица представляет собой массив синаптических соединений нейронов, и выход из строя значительного числа синаптических связей нейросети слабо отражается на ее работе [4].

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 ее 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 [5].

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 [6].

Использование кроссбара из комплементарных мемристоров в запоминающей матрице впервые представлено в [7]. 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 [8–10] эта матрица, позволяющая взвешивать и суммировать входные сигналы в виде постоянных напряжений, применена в качестве аппаратной реализации персептрона. Она же использовалась в [11], где на основе биоморфного подхода выполнялось суммирование входных импульсов напряжения.

Однако разработанная в [8–11] матрица не может быть использована в качестве сверхбольшой запоминающей матрицы нейропроцессора из-за низкой энергоэффективности при записи и высокой деградации выходного сигнала при считывании. Для построения большой архитектуры нейропроцессора эту матрицу можно было бы использовать как основу отдельных кластеров сверхбольшой матрицы, но дополнительные схемы коммутации между кластерами приведут к значительному уменьшению интеграции элементов.

Проблема энергоэффективности сверхбольшой запоминающей матрицы решается путем использования мемристорно-диодного кроссбара, ячейка которого представляет собой двухслойное соединение комплементарных мемристоров и одного разделяющего диода Зенера. Кроме того, применение диода Зенера позволяет уменьшить деградацию выходного сигнала при суммировании входных импульсов напряжения.

## 3.1. ПЛАНАРНАЯ ДВУХСЛОЙНАЯ ЗАПОМИНАЮЩАЯ МАТРИЦА НА ОСНОВЕ ИНТЕГРАЦИИ ЭЛЕМЕНТАРНЫХ ЯЧЕЕК

В разд. 1.5 представлен новый компонент наноэлектроники — мемристорно-диодный кроссбар [12], необходимый для создания сверхбольшой запоминающей матрицы биоморфного нейропроцессора.

На основе двух топологий комплементарной мемристорно-диодной ячейки разработаны два варианта запоминающей матрицы — с параллельным и последовательным выводом данных (рис. 3.1) [13].

Линии *X* и *Y* на рис. 3.1, *б*, 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 [14]. 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 [15]. 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 все время удерживается высоким, а сквозной ток через комплементарные мемристоры остается минимальным, что повышает энергоэффективность всей матрицы. Подача импульса тока осуществляется через прямосмещенный 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.

Результат записи в первую ячейку, полученный в ходе SPICE-моделирования работы матрицы из 2 × 2 ячеек, представлен на рис. 3.2, 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 $M_1$ (синяя кривая) увеличивается

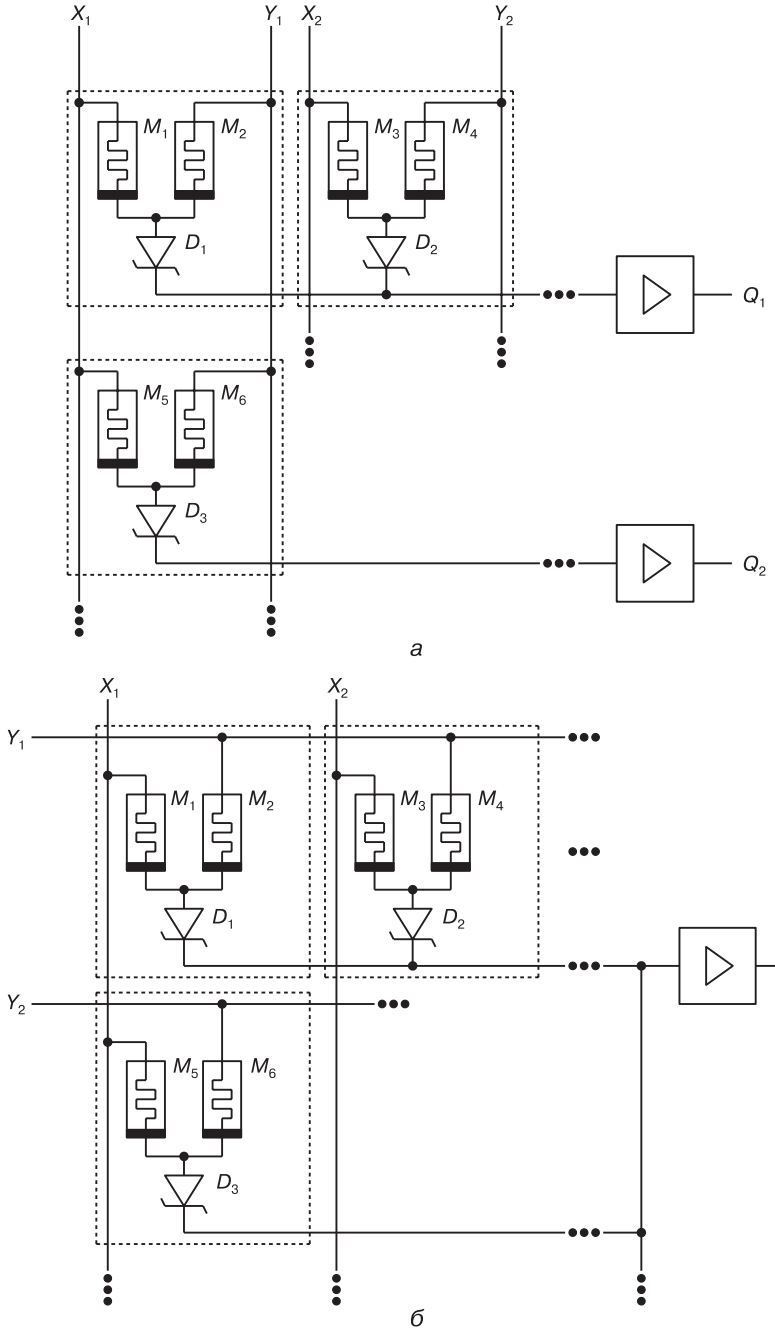



**Рис. 3.1.** Матрица запоминающего устройства:

*а* — с параллельным выводом данных через общую шину;
*б* — с последовательным выводом данных через общую шину

до предела (1 МОм), а второго $M_2$ (оранжевая кривая) уменьшается до минимального (10 кОм). Таким образом, общее сопротивление пары остается высоким.

Моделирование работы квадратных матриц с разным числом ячеек показывает, что затухание амплитуды напряжения сигнала до порогового значения 1 В происходит в матрице из $N = 2{,}5 \cdot 10^4$ ячеек. На рис. 3.3 точками показаны значения напряжения в первых 100 ячейках, а прямая линия соответствует аппроксимации методом наименьших квадратов.

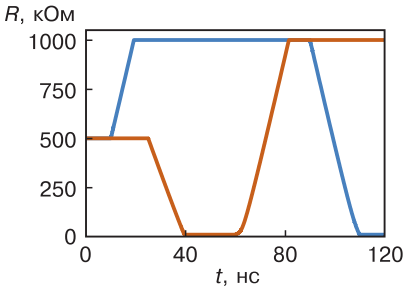



**Рис. 3.2.** Кривая изменения сопротивлений $M_1$ (синий) и $M_2$ (оранжевый) от времени

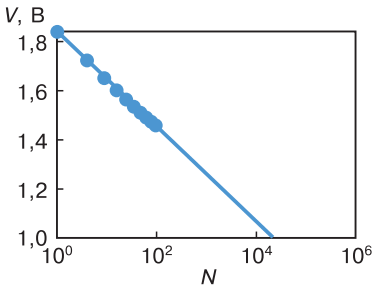



**Рис. 3.3.** Затухание амплитуды напряжения сигнала от числа ячеек в матрице

Отличающийся более высокой скоростью принцип переключения комплементарных ячеек без участия диода Зенера возможен только для матрицы с побитным доступом, показанной на рис. 3.1, *б*. В этом случае, с помощью драйверов управляющих сигналов подается 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.

Проблема энергоэффективности сверхбольшой запоминающей матрицы нейропроцессора решается путем использования комплементарной мемристорно-диодной ячейки, которая представляет собой двухслойное соединение комплементарных мемристоров и одного разделяющего диода Зенера. Кроме того, применение диода Зенера позволяет уменьшить деградацию выходного сигнала при суммировании входных импульсов напряжения.

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 остается высоким. При этом через ячейки протекает минимальный ток, что и определяет энергоэффективность матрицы.

Получен патент [16] на планарные сверхбольшие запоминающие матрицы с энергонезависимой памятью, высокой степенью интеграции элементов и малым энергопотреблением, обеспечивающие параллельный и последовательный доступ к записи и считыванию данных.

## 3.2. ЭЛЕКТРИЧЕСКАЯ СХЕМА, ТОПОЛОГИЯ И НАНОТЕХНОЛОГИЯ ИЗГОТОВЛЕНИЯ 3D ЗАПОМИНАЮЩЕЙ МАТРИЦЫ

Дальнейшее увеличение интеграции элементов может быть достигнуто путем объединения на кристалле планарных двухслойных запоминающих матриц в 3D-структуру.

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 [3]. 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.

В [17] представлена разработка электрической схемы, топологии и технологии изготовления 3D сверхбольшой запоминающей матрицы на основе комбинированного мемристорно-диодного кроссбара, в которой помимо хранения информации в виде веса синаптической связи происходит обработка входных данных согласно принципам нейроморфного подхода — сначала взвешивание напряжения входного сигнала при прохождении через мемристор, а затем — суммирование взвешенных сигналов в сумматоре (например, в конденсаторе или мемристоре).

Высокая степень интеграции запоминающей матрицы достигается за счет трехмерной компоновки элементов, образующих 3D-структуру. Структура 3D состоит из одинаковых горизонтально расположенных и зеркально ориентированных по отношению друг к другу комбинированных кроссбаров, включающих ячейки 1D2M. Соединение между кроссбарами осуществляется по общим шинам строк и столбцов. Два соседних кроссбара

с общими шинами, соединенными с катодами диодов Зенера, образует отдельный функциональный пласт. Все функциональные пласты идентичны, накладываются друг на друга и объединяются общими шинами в электрическую цепь. Принцип объединения комбинированных кроссбаров в электрическую цепь показан на рис. 3.4 [18]. Электроды ячеек $Y_1$ и $Y_2$ 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 $X_1$–$X_4$. Два соседних перекрестья металлических проводников образуют комплементарные мемристоры.

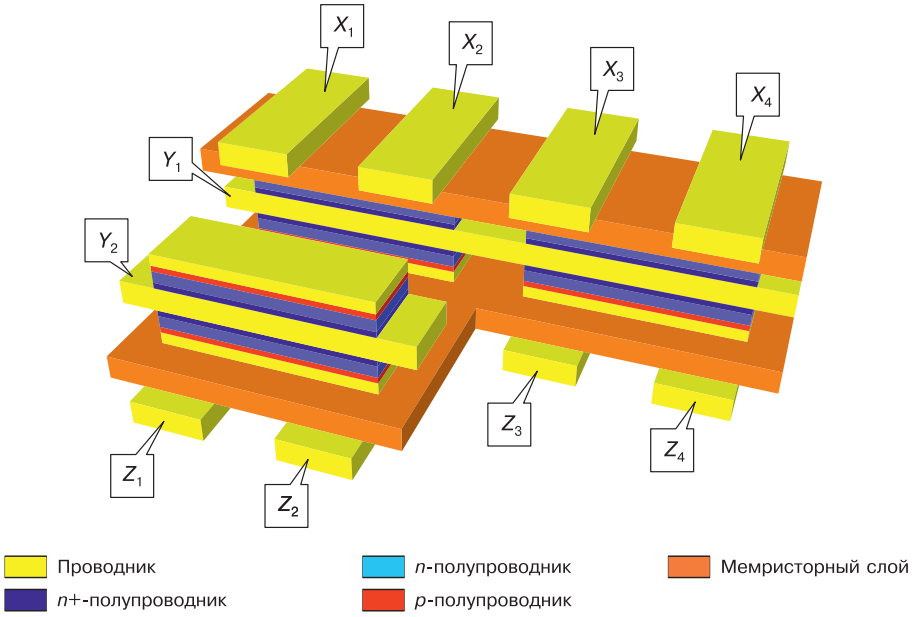



**Рис. 3.4.** Принцип объединения комбинированных кроссбаров в 3D запоминающую матрицу с высокой интеграцией элементов

На рис. 3.5 представлена топология запоминающей матрицы, состоящей из трех горизонтально расположенных и зеркально ориентированных по отношению друг к другу кроссбаров, в каждом из которых находится 18 ячеек.

Электрическая схема фрагмента трехмерной запоминающей матрицы из трех комбинированных кроссбаров, поясняющая принцип соединения соседних пластов, приведена на рис. 3.6.

Получен патент на электрическую схему и топологию 3D сверхбольшой запоминающей матрицы на основе комбинированного мемристорно-диодного кроссбара [19].

Предлагаемая 3D запоминающая матрица работает в двух режимах: режиме записи и режиме чтения данных.

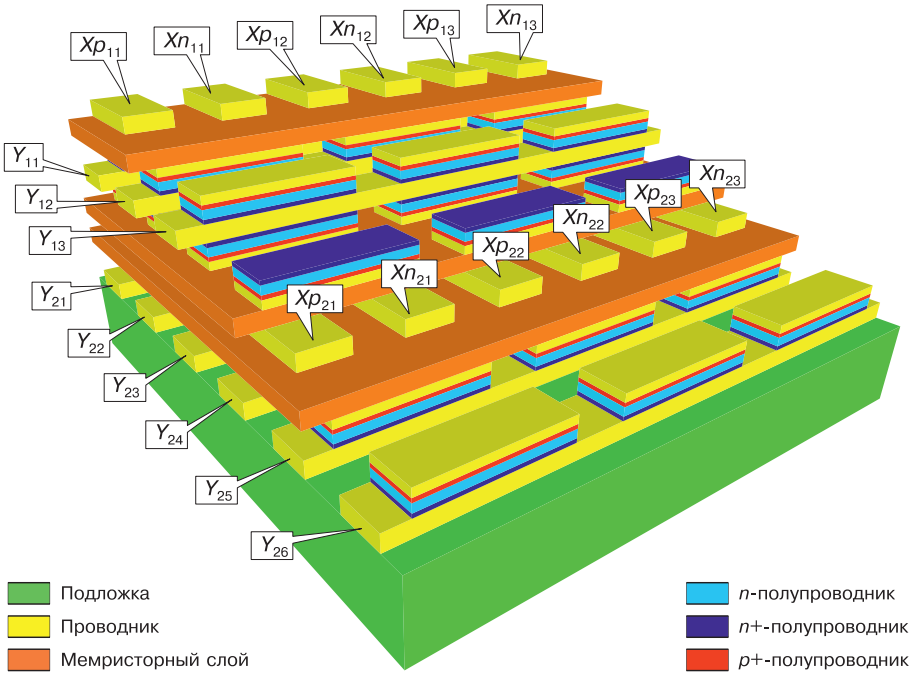



**Рис. 3.5.** Топология 3D запоминающей матрицы

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 $M_1$ и $M_2$, это напряжение поступает с одной стороны через проводящие шины $Xp_{11}$ и $Xn_{11}$, являющиеся контактами мемристоров. С другой стороны напряжение подается через диод Зенера $D_1$ по выходной шине $Y_{21}$ 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.

Сверхбольшая запоминающая матрица изготавливается согласно представленной топологии с помощью вакуумной магнетронной технологии. Сначала на подложку через маску наносятся проводники $Y_{21}$—$Y_{26}$ путем распыления металлического катода.

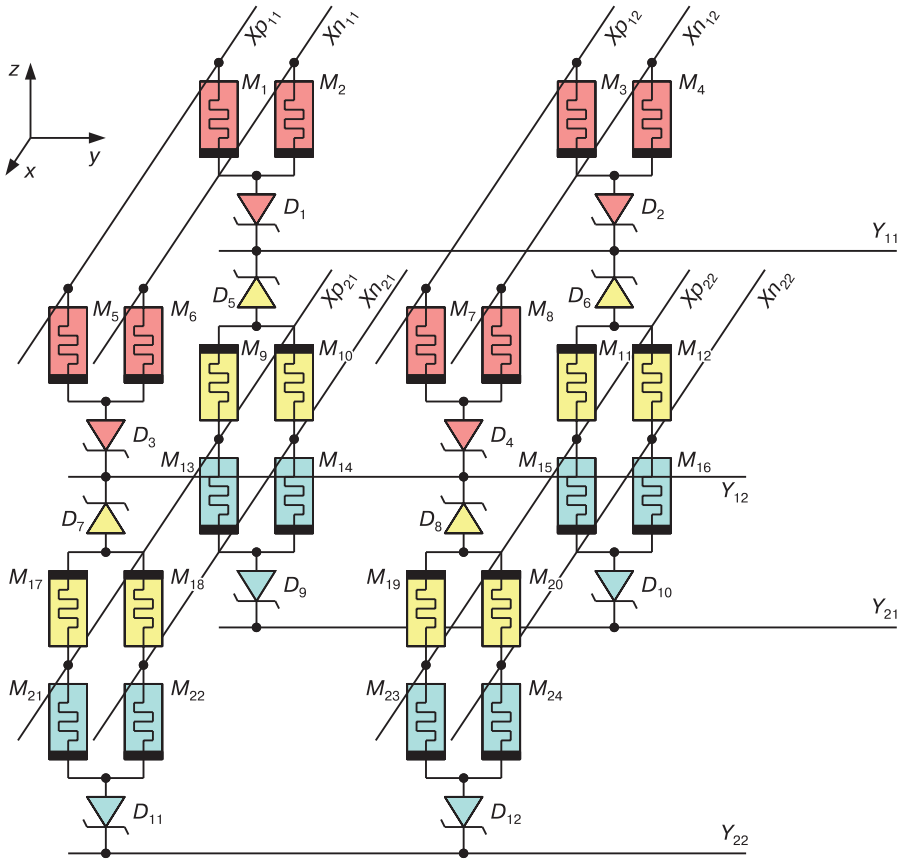



**Рис. 3.6.** Электрическая схема фрагмента трехмерной запоминающей матрицы. Три цвета соответствуют трем комбинированным кроссбарам

Затем пустое пространство между проводниками методом реактивного магнетронного распыления заполняется изолятором (например, диоксидом кремния). Далее через маску (или с помощью сухого травления) последовательно формируются три полупроводниковых слоя диодов Зенера с разной проводимостью в результате одновременного распыления в магнетроне катодов кремния и легирующей примеси. На диоды Зенера через ту же маску наносится общий металлический электрод будущих комплементарных мемристоров. Свободное пространство между полученными структурами также заполняется диэлектриком. Затем реактивным магнетронным распылением наносится пленка оксида переходного металла (например, $TiO_2$), являющаяся мемристорным слоем. Далее на эту пленку через маску наносятся проводники столбцов $Xp_{21}$–$Xp_{23}$, $Xn_{21}$–$Xn_{23}$ ортогонально шинам строк $Y_{21}$–$Y_{26}$ так, чтобы над каждым диодом проходило два проводника. Таким образом,

на кристалле образуется один функциональный пласт, состоящий из диодного и мемристорного слоев с промежуточными слоями проводников.

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.

Вакуумная технология нанесения мемристорного слоя и легированных полупроводниковых слоев диода Зенера в магнетронном технологическом модуле с двумя одновременно распыляющимися катодами является достаточно простой и позволяет изготавливать сверхбольшие запоминающие матрицы, которые в первую очередь могут быть использованы в нейропроцессорах. Достоинством такой технологии является то, что изготовление комбинированного кроссбара осуществляется в одном технологическом модуле, причем изготовление диода Зенера по этой технологии гораздо проще термодиффузионного легирования [20] и имплантации примесей в полупроводник [21].

## 3.3. ВЗВЕШИВАНИЕ НАПРЯЖЕНИЙ ВХОДНЫХ СИГНАЛОВ И СУММИРОВАНИЕ ВЫХОДНЫХ НАПРЯЖЕНИЙ И ТОКОВ ЯЧЕЕК

Рассмотрим работу фрагмента электрической схемы матрицы [17]. Математическая функция, выполняемая формальным нейроном, имеет вид:

$$y = f_{\text{act}}\left(\sum_i w_i x_i\right), \tag{3.1}$$

где $f_{\text{act}}$ — функция активации, а в ее аргументе $x_i$ — сигналы от предыдущих нейронов; $w_i$ — веса нейроных связей.

Основной процесс в запоминающей матрице — сложение и вычитание взвешенных сигналов — выполняется путем умножения напряжения входного сигнала на функцию сопротивления с помощью закона Ома и суммирования получившихся токов по первому закону Кирхгофа. Входной импульс подается на контакты выбранной ячейки в виде двух импульсов напряжения противоположной полярности, абсолютная величина которых меньше порогового напряжения переключения мемристора. В результате комплементарная пара мемристоров образует резистивный делитель напряжения.

Напряжение на входе сумматора при приложении к контактам мемристоров ячейки напряжения разной полярности амплитудой $U_{\text{input}}$ имеет вид

$$U_i = \left(\frac{2U_{\text{input}}}{R_{1i} + R_{2i}} R_{2i} - U_{\text{input}}\right)\frac{1}{R_{\text{D}} + R_{\text{p}}} R_{\text{p}}, \tag{3.2}$$

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 сопротивления остается постоянной: $R_{1i} + R_{2i} = K = \text{const}$. Учитывая, что сопротивление открытого диода мало, $R_D \ll R_p$, то справедливы следующие выражения для напряжений на входе и выходе сумматора соответственно:

$$U_i = \frac{U_{\text{input}}}{K}\left(2R_{1i} - K\right); \tag{3.3}$$

$$U_S = \sum_i U_i = \frac{U_{\text{input}}}{K}\sum_i \left(2R_{1i} - K\right). \tag{3.4}$$

Суммирование напряжений выполняется последовательной подачей напряжения на выбранные ячейки. При подключении выходной линии кроссбара к усилителю с высокоомным входом и нагрузкой в виде конденсатора будут складываться выходные напряжения ячеек.

Таким образом, соотношение электрических величин в аппаратной части, описываемое выражением (3.4), соответствует аргументу активационной функции в формуле (3.1) следующим образом

$$x_i \to U_{\text{input}}; \quad w_i \to \frac{2R_{1i} - K}{K}.$$

Сигнал $U_{\text{input}}$ 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. Взаимовлияния ячеек не будет, поскольку входные инверторы неиспользуемых ячеек находятся в высокоимпедансном состоянии.

Сложение токов, протекающих через закрытые диоды Зенера, возможно при низком входном сопротивлении усилителя. Взвешивание напряжений входных импульсов с последующим суммированием токов позволяет уменьшить паразитные токи между ячейками и произвести корректное сложение одновременно пришедших импульсов [22]. На рис. 3.7 показан выходной ток одной шины матрицы как результат сложения токов из двух ячеек, находящихся в разных синаптических состояниях.

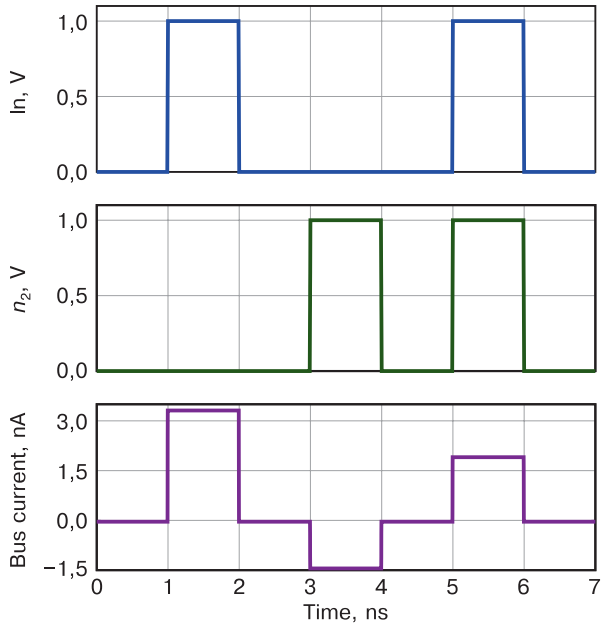



**Рис. 3.7.** Сложение выходных токов двух ячеек запоминающей матрицы на общей выходной шине

Первый импульс подавался на первую ячейку с мемристорами $M_1$ и $M_2$, второй импульс — на вторую ячейку с мемристорами $M_3$ и $M_4$. Из рис. 3.6 видно, что протекающий через выходную шину $y_{11}$ ток при одновременной подаче импульсов на обе ячейки равен сумме токов от отдельных ячеек.

## 3.4. ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ РАБОТЫ ЗАПОМИНАЮЩЕЙ МАТРИЦЫ

Для моделирования работы запоминающей матрицы в режимах записи (см. разд. 1.7), взвешивания и суммирования сигналов использовалась оригинальная специализированная программа MDC-SPICE, разработанная для расчета больших электрических схем с мемристорно-диодными кроссбарами. Моделирование, в частности, показало, что производительность матрицы слабо зависит от ее размера. Под производительностью имеется в виду задержка между подачей входных сигналов в запоминающую матрицу и срабатыванием логических вентилей, вызванным изменением напряжений на выходных проводниках матрицы. В матрице размером 50 × 50 с 90 нм инверторами задержка сигнала, обусловленная преимущественно входной емкостью транзисторов инвертора, составляет 5,62 нс. Она соответствует отрезку времени от подачи на вход матрицы идеального прямоугольного импульса напряжения до изменения напряжения на выходе инвертора.

Работу с отдельными входными импульсами можно рассматривать как последовательное считывание. Схема подачи входных импульсов описана в разд. 3.2. Из результатов моделирования следует (рис. 3.8), что существенная деградация напряжения импульса, с 1 В на входе до 0,3 В на выходе, происходит в матрице размером 1000 × 1000 ячеек при отношении максимального и минимального сопротивлений мемристора $R$, равным 100. Если в одной миниколонке неокортекса в среднем находится $10^2$ нейронов, то запоминающая матрица такого объема может имитировать работу пятой части гиперколонки.

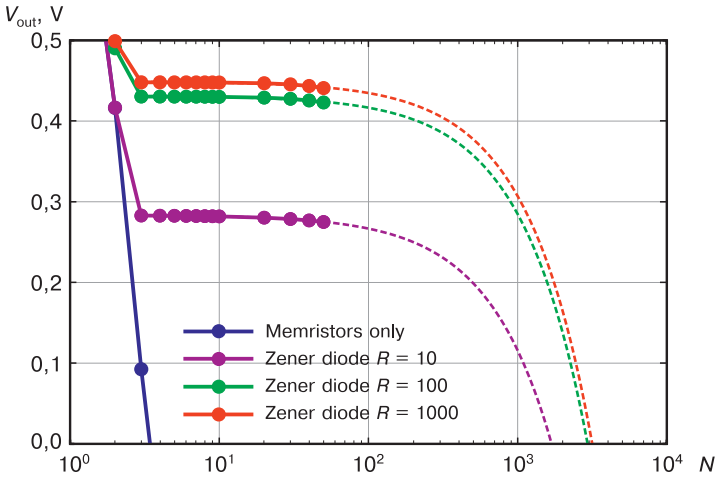



**Рис. 3.8.** Зависимость выходного напряжения $V_{out}$ от размера квадратной матрицы $N \times N$ при подаче одного импульса амплитудой 1 В и последовательном чтении для разных $R$

Из рис. 3.8 видно, что при отсутствии в ячейках диода Зенера [11] выходное напряжение снижается практически до нуля уже в 3 × 3 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 ~0,3 В достаточен для выполнения дальнейшей процедуры суммирования.

Схема ввода входных импульсов в матрицу комплементарных мемристоров без диодов, предложенная в [13], 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. 3.9 представлены результаты SPICE-моделирования процесса прохождения одного импульса с амплитудой 1 В (считывания состояния одной ячейки) по этой схеме.

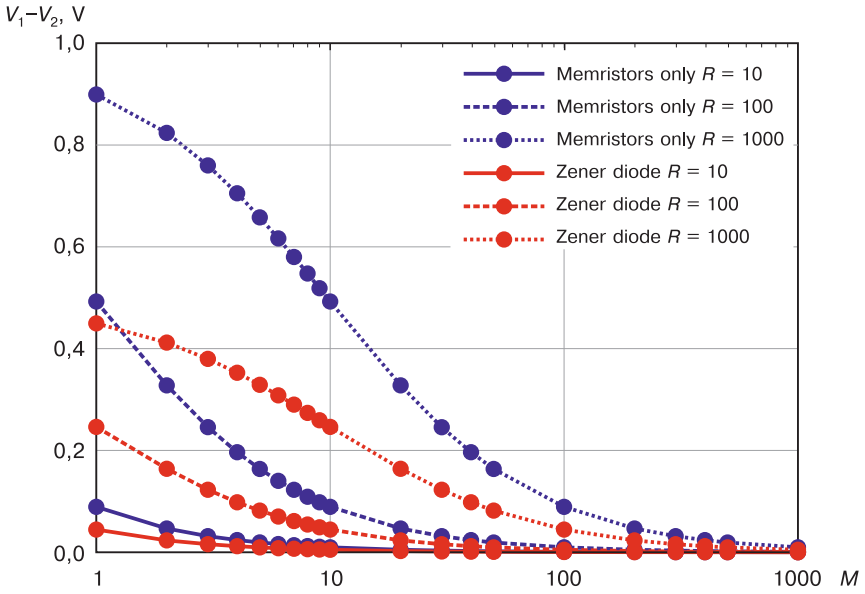



**Рис. 3.9.** Зависимость выходного дифференциального напряжения в зависимости от размера матрицы при последовательном чтении при разных $R$

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 [13] не может быть использована в качестве сверхбольшой матрицы нейропроцессора.

Далее приводятся результаты SPICE-моделирования режима суммирования импульсов в предложенной сверхбольшой матрице [17].

В качестве сумматора взята простая схема из КМОП-инвертора и конденсатора (рис. 3.10). Сильная нелинейность передаточной характеристики инвертора, образованного транзисторами $T_1$ и $T_2$, 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.

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. 3.11 показано изменение во времени выходного напряжения сумматора $V_{out}$.

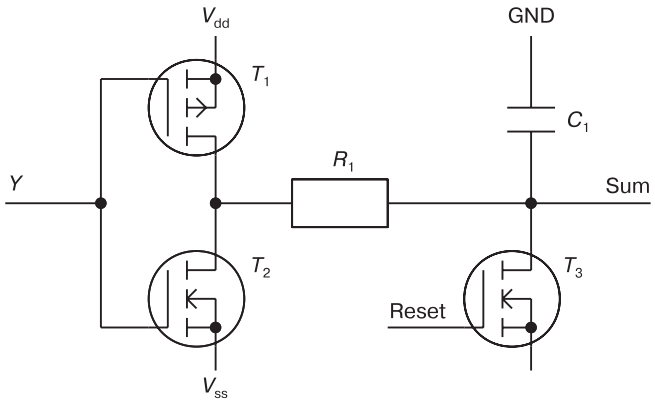



**Рис. 3.10.** Схема сумматора:
$V_{dd}$ и $V_{ss}$ — полюса питания; GND — шина с нулевым потенциалом

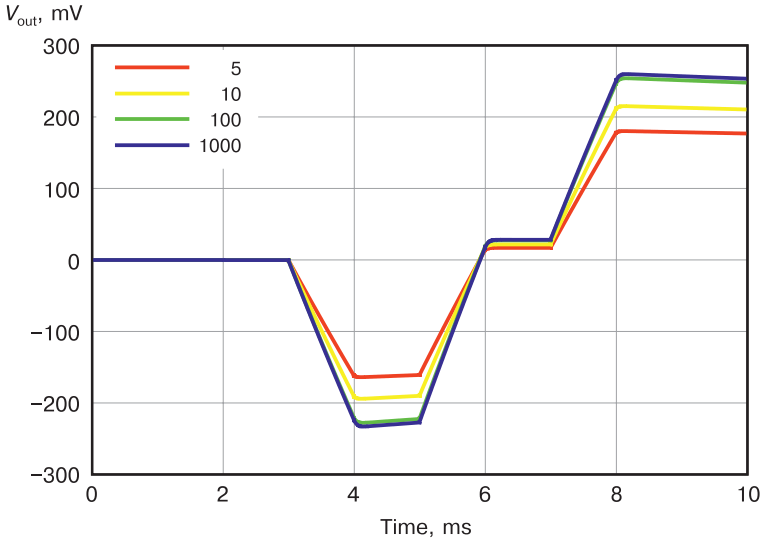



**Рис. 3.11.** Выходное напряжение сумматора.
Цвет соответствует различным *R*

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 матрице с размером матрицы от 1 × 1 до 50 × 50 без учета емкости усилителя время нарастания на выходной шине практически постоянно и составляет около 41 пс. При моделировании

учитывалась емкость проводников (толщиной 30 нм и с межшинным расстоянием 200 нм), диодов (площадь $600 \times 200$ нм$^2$) и мемристоров (площадь $200 \times 200$ нм$^2$, толщина 30 нм). Моделирование матрицы с моделью инвертора PTM 90 нм [23] дает задержку 5,62 нс.

## 3.5. АССОЦИАТИВНОЕ САМООБУЧЕНИЕ СИНАПСОВ ЗАПОМИНАЮЩЕЙ МАТРИЦЫ И ГЕНЕРАЦИЯ НОВОЙ АССОЦИАЦИИ

Ассоциативное самообучение и формирование новой ассоциации в нейросети с мемристорными синапсами проводится по правилу Хебба [24]. В этом параграфе с помощью SPICE-моделирования процесса ассоциативного самообучения продемонстрирована генерация новой ассоциации (нового знания) в мемристорно-диодном кроссбаре в отличие от существующих нейросетей с синапсами на базе дискретных мемристоров [25–29].

Суть ассоциативного самообучения запоминающей матрицы заключается в особом процессе, когда биоморфные информационные импульсы, формируемые выходом возбужденного нейрона в нейросети, приводят к усилению его синаптических связей. Усиление связей возбужденного нейрона с нейронами предыдущего слоя происходит, если в этот момент времени эти нейроны тоже оказались возбужденными и сформировали на своих выходах информационные импульсы. Процесс ассоциативного обучения является биоморфным и принципиально схожим с известным экспериментом русского физиолога И.П. Павлова [29] по формированию условного рефлекса у собаки. Его можно использовать для проверки формирования новой ассоциации в нейропроцессоре.

Предполагается, что обучение синапсов в запоминающей матрице будет происходить по правилу Хебба, как и в реальном синапсе: сила связи между одновременно активировавшимися нейронами увеличивается. Для реализации индуцированной долговременной потенциации синапса была выбрана схема из трех нейронов (рис. 3.12). Согласно схеме выходные импульсы двух нейронов через синаптические связи, представленные комплементарными мемристорно-диодными ячейками, поступают на вход третьего нейрона.

SPICE-моделирование синаптической пластичности проводилось в электрической схеме нейрона (см. рис. 1.19) при ассоциации выхода нейрона $n_2$ со входом нейрона $n_3$. Параметры мемристоров и диодов Зенера такие же, как в [17]: сопротивления мемристоров в высокопроводящем и низкопроводящем состояниях равны 1 кОм и 100 кОм, напряжения пробоя и прямого смещения диода Зенера 1,2 В и 0,3 В соответственно. Напряжение открытия диодов 0,3 В, коэффициент преобразования ИТУТ $3 \cdot 10^4$, емкость конденсатора 1 пФ, порог генерации выходного импульса 100 мВ. Амплитуды сигнального и программирующего напряжения равны 1 и 5 В соответственно.

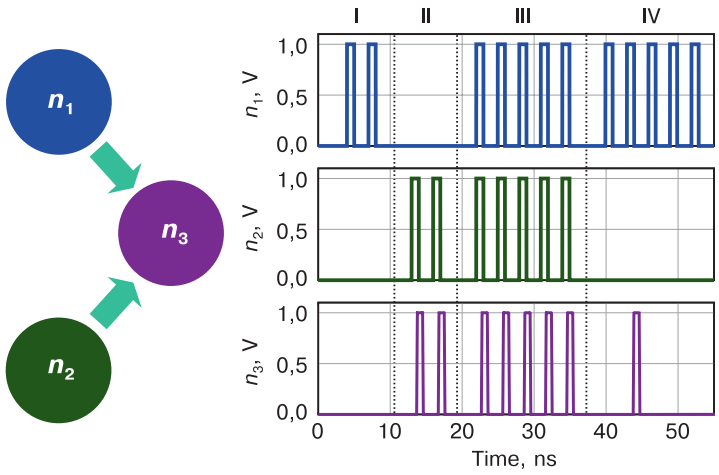



**Рис. 3.12.** Возникновение новой ассоциации в группе из трех нейронов. Импульсы напряжения, генерируемые нейронами на разных этапах ассоциативного обучения

На рис. 3.12 показаны выходные сигналы нейронов, полученные в результате моделирования. На первом этапе обучения двух импульсов нейрона $n_1$ недостаточно для активации нейрона $n_3$, так как связь между ними слабая. С другой стороны, нейрон $n_2$ каждым своим импульсом способен активировать нейрон $n_3$. На третьем этапе происходит потенциация связи между нейроном $n_1$ и нейроном $n_2$. В результате усиления связи двух импульсов от нейрона $n_1$ достаточно для активации нейрона $n_3$. Потенциация синапса между нейроном $n_1$ и нейроном $n_3$ будет происходить из-за того, что они активируются одновременно, и напряжение на мемристорах, соответствующих этой связи, будет выше порогового. Таким образом, продемонстрирована генерация новой ассоциации (нового знания) в группе из трех нейронов.

Складывая мощность, выделяемую всеми источниками напряжения электрической схемы, получим ее полное энергопотребление во времени (рис. 3.13).

Из рис. 3.13 видно, что энергопотребление в нейропроцессоре при обучении резко повышается при наличии потенциалов действия с 8,1 нВт до нескольких десятков микроватт. Высокие короткие всплески энергопотребления соответствуют перекрытию по времени импульсов от второго и третьего нейронов, показанных на рис. 3.12. Средняя потребляемая электрической схемой трех нейронов мощность за время моделирования составила 107 мкВт.

Возможность переобучения нейросети реализована безусловной релаксацией состояния синапсов к начальному состоянию путем периодической подачи надпороговых разобучающих импульсов с отдельного генератора. Этот эффект постепенно ослабляет синапс между нейроном $n_1$ и нейроном $n_3$. На последнем этапе IV активация нейрона $n_3$ импульсами от $n_1$ происходит все реже (см. рис. 3.12). Изменение состояния синапса

во времени определяется суммарным действием двух механизмов: условного обучения по правилу Хебба и безусловного разобучения — забывания. Соответственно, чувствительность ассоциативного обучения можно настраивать путем изменения частоты генератора разобучающих импульсов $f_f$. На рис. 3.14 представлено изменение проводимости $\Delta\sigma$ одного из мемристоров синапса в зависимости от соотношения $F = f_e/f_f$ при одинаковой ширине импульсов, где $f_e$ — частота программирующих импульсов.

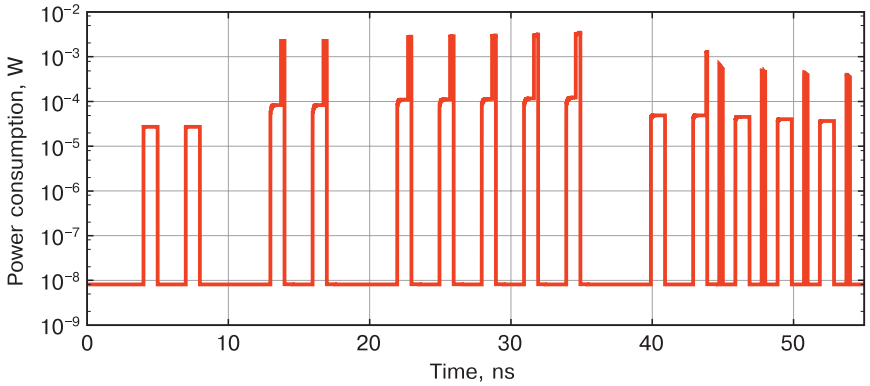



**Рис. 3.13.** Изменение во времени потребляемой мощности схемы из трех нейронов при обучении синапсов запоминающей матрицы

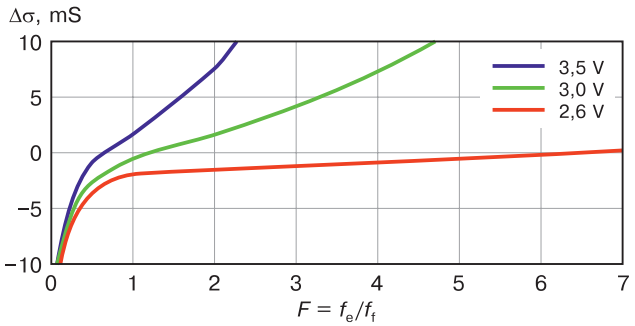



**Рис. 3.14.** Чувствительность матрицы к обучению

Из рис. 3.14 следует, что с увеличением величины напряжения программирования мемристоров возрастает скорость изменения силы синапса.

В отличие от электрических схем самообучения мемристорных синапсов в [25; 26; 28], включающих несколько операционных усилителей, предложенная электрическая схема нейрона (см. рис. 1.19) содержит только один элемент, работающий в активном режиме (источник тока). Поэтому она может быть использована для построения сверхбольшой аппаратной нейросети с высокими интеграцией элементов и энергоэффективностью.

Список литературы


1. *Chevallier C.J.*, *Siau C.H.*, *Lim S.F.* et al. A 0.13 μm 64 Mb multi-layered conductive metal-oxide memory // Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International. 2010. Vol. 1. Pp. 260–261.
2. *Liu T.*, *Yan T.H.*, *Scheuerlein R.* et al. 130.7 $mm^2$ 2-Layer 32-Gb ReRAM memory device in 24-nm technology // IEEE Journal of Solid-State Circuits. 2014. Vol. 49. No. 1. Pp. 140–153.
3. *Bandyopadhyay A.*, *Scheuerlein R.E.*, *Gorla C.R.*, *Le B.* FET low current 3D ReRAM non-volatile storage // 2015. US Patent No. 0070965 A1.
4. *Haykin S.* Neural networks: A comprehensive foundation. Second ed. // Prentice-Hall. Upper Saddle River. NJ. USA. 1999.
5. *Bobylev A.N.*, *Udovichenko S.Yu.* The electrical properties of memristor devices $TiN/Ti_xAl_{1-x}O_y/TiN$ produced by magnetron sputtering // Russian Microelectronics. 2016. Vol. 45. No. 6. Pp. 396–401.
6. *Xia O.*, *Robinett W.*, *Cumbie M.* et al. Memristor-CMOS hybrid integrated circuits for configurable logic // Nano Letters. 2009. Vol. 9. No. 10. Pp. 3640–3645.
7. *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. Vol. 74. No. 6. Pp. 2484–2496.
8. *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. Vol. 10. No. 1, 5.
9. *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. Vol. 11. No. 4. P. 34.
10. *Chabi D.*, *Zhaohao W.*, *Bennet C.* et al. Ultra high density memristor neural crossbar for on-chip supervised learning // IEEE Transactions on Nanotechnology. 2015. Vol. 14. No. 6. Pp. 954–962.
11. *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. Pp. 125–130.
12. *Писарев А.Д.*, *Бусыгин А.Н.*, *Бобылев А.Н.*, *Удовиченко С.Ю.* Комбинированный мемристорно-диодный кроссбар как основа запоминающего устройства // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2017. № 4. С. 142–149.
13. *Maevsky O.V.*, *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.* Complementary memristor-diode cell for a memory matrix in neuromorphic processor // International journal of nanotechnology. 2018. Vol. 15. No. 4/5. Pp. 388–393.
14. *Keeth B.*, *Baker R.*, *Johnson B.*, *Lin F.* DRAM Circuit design: Fundamental and high-speed topics. Second ed. // Wiley-IEEE Press. 2007. P. 440.
15. *Baker R.* CMOS: Circuit design, layout, and simulation. 3rd ed. // Wiley-IEEE Press. 2010. P. 1208.
16. *Маевский О.В.*, *Писарев А.Д.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* Запоминающее устройство на основе комплементарной мемристорно-диодной ячейки. 2018. Патент № 2649657.
17. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Yu.*, *Maevsky O.V.* 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor // Microelectronic Engineering. 2018. Vol. 198. Pp. 1–7.

18. *Писарев А.Д.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* и др. 3D запоминающая матрица на основе комплементарной мемристорно-диодной ячейки. 2019. Патент №2697623.
19. *Vinet M.*, *Batude P.*, *Tabone C.* et al. 3D monolithic integration: Technological challenges and electrical results // Microelectronic Engineering. 2011. V. 88. Pp. 331–335.
20. *Chen W.*, *Lin X.*, *Parris P.M.* Zener diode device and fabrication. 2014. Patent US 0061715 A1.
21. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor-diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
22. *Zhao W.*, *Cao Y.* Predictive technology model for nano-CMOS design exploration // ACM Journal on Emerging Technologies in Computing Systems. 2007. Vol. 3. No. 1. P. 1.
23. *Pershin Y.V.*, *Di Ventra M.* Experimental Demonstration of Associative Memory with Memristive Neural Networks // Neural Networks. 2010. Vol. 23. No. 7. Pp. 881–886.
24. *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. Vol. 65. No. 7. Pp. 2210–2220.
25. *Yang L.*, *Zeng Z.*, *Huang Y.*, *Wen S.* Memristor-based circuit implementations of recognition network and recall network with forgetting stages // IEEE Transactions on Cognitive and Developmental Systems. 2018. Vol. 10. No. 4. Pp. 1133–1142.
26. *Wang Z.*, *Rao M.*, *Han J.-W.* et al. Capacitive neural network with neuro-transistors // Nature Communications. 2018. Vol. 9. P. 3208.
27. *Zhang X.*, *Long K.* Improved learning experience memristor model and application as neural network synapse // IEEE Access. 2019. Vol. 7. Pp. 15262–15271.
28. *Minnekhanov A.A.*, *Emelyanov A.V.*, *Lapkin D.A.* et al. Parylene based memristive devices with multilevel resistive switching for neuromorphic applications // Scientific Reports. 2019. Vol. 9. P. 10800.
29. *Pavlov I.P.* Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex // Annals of Neurosciences. 2010. Vol. 17. No. 3. Pp. 136–141.

# ГЛАВА 4

# УНИВЕРСАЛЬНАЯ ЛОГИЧЕСКАЯ МАТРИЦА НА ОСНОВЕ КОМБИНИРОВАННОГО МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА

В настоящее время разработано несколько логических схем с применением мемристоров [1–10]. В схемах Memristor-Based Material Implication (IMPLY) [4–7; 10] и Memristor-Aided Logic (MAGIC) [8] в качестве логических значений использованы не уровни напряжения, а сопротивление мемристоров. Это приводит к дополнительным операциям преобразования переменных при работе совместно с традиционными логическими схемами. Такого недостатка лишена схема Memristor Ratioed Logic (MRL) [9], в которой входные и выходные сигналы представлены в виде напряжений. Однако для вычисления результата логической операции в схемах [4–10] выполняется несколько переключений мемристоров. Современная двоичная логика предполагает срабатывание логических элементов (например, КМОП-вентилей) до нескольких миллиардов раз в секунду, в то время как ресурс переключения у мемристоров достаточно мал. Поэтому нецелесообразно использование логических вентилей, в которых при выполнении логической операции происходит переключение мемристоров. Мемристоры следует использовать там, где число переключений относительно мало.

Мемристорный массив Акерса [1], сотоящий из ячеек $2M$, является примером логического устройства, в котором в процессе вычисления логических функций переключение мемристоров не происходит. Элементарная ячейка этого массива является делителем напряжения из двух мемристоров, включенных антипоследовательно. Из-за большого числа пассивных ячеек, включенных последовательно, массив Акерса обладает высокой деградацией выходного напряжения. Кроме этого, для обеспечения возможности независимой записи каждая ячейка массива содержит четыре транзистора, что ведет к низкой интеграции.

Другой вариант применения мемристоров в логических схемах заключается в построении маршрутизатора логических сигналов в программируемой логической схеме. В [2; 3] предложены двухслойные программируемые логические устройства, в которых использованы мемристорные коммутирующие ячейки, причем в [3] представлена техническая реализация. В этих устройствах мемристорный кроссбар расположен под углом

над КМОП-вентилями. Организация ячеек в кроссбар без селективных элементов обеспечивает большую плотность элементов, но обладает существенным недостатком, связанным с протеканием паразитных токов через соседние ячейки. По этой причине организация больших кроссбаров невозможна без дополнительных элементов. Для исключения этого эффекта достаточно последовательно мемристору включить селективный элемент (например, диод или *n*-МОП-транзистор) [11]. В свою очередь применение диода исключает возможность перепрограммирования биполярных мемристоров. Несмотря на то что *n*-МОП-транзистор может пропускать ток в обоих направлениях, один из режимов является неоптимальным. Перспективно использование диода Зенера, пропускающего ток в обоих направлениях.

Возможно использование мемристоров для выполнения математических расчетов в аналоговом режиме, а не только в качестве ключей для коммутации логических элементов. В этом случае обработка сигнала достигается путем умножения матрицы на вектор с помощью кроссбара, элементарная ячейка которого состоит из мемристора и транзистора (1T1M архитектура) [12]. Представленный массив назван большим и содержит 8192 ячейки. При некоторой модификации, заключающейся в ином способе подключения периферийных устройств, можно из этого массива получить логический массив, а ячейку 1T1M рассматривать как базовую для построения большого логического блока.

Однако перечисленные логические матрицы не могут быть использованы в качестве сверхбольшой запоминающей матрицы нейропроцессора из-за низких интеграции элементов и энергоэффективности.

## 4.1. ПЛАНАРНАЯ ДВУХСЛОЙНАЯ ЛОГИЧЕСКАЯ МАТРИЦА НА ОСНОВЕ ИНТЕГРАЦИИ ЭЛЕМЕНТАРНЫХ ЯЧЕЕК

### 4.1.1. Мемристорная ячейка с транзисторами для блока логического коммутатора

В работе [13] дана электрическая схема коммутационной ячейки, состоящей из транзисторной КМОП-структуры и одного мемристора (рис. 4.1, *а*). Мемристор первым контактом подключен к соединению стоков *n*- и *p*-канальных транзисторов, которые переключают этот контакт между полюсами источника питания $V_{ss}$ и $V_{dd}$.

Топология матрицы представляет собой структуру из двух основных слоев: нижний КМОП-слой транзисторов, на который через изолятор нанесен верхний слой мемристоров. Технологические этапы создания областей ячейки показаны срезами на рис. 2, *б*, которые условно помечены буквами *A*–*B*–*C*–*D*–*E*. Проводники и области каждого типа проводимости на рис. 4.1, *б* отмечены разными цветами. На этапах *A*–*B*–*C* формируют

слой КМОП-транзисторов по стандартной технологии: на этапе *A* на подложке создаются карманы двух полевых *n*- и *p*-канальных транзисторов; на этапе *B* формируются затворы транзисторов; на этапе *C* транзисторы соединяют проводниками в комплементарные пары.

Далее на заготовке, например, методом магнетронного напыления и с помощью литографии формируют слой диэлектрика с матрицей проводящих переходных колодцев, которые присоединяют стоки комплементарных пар транзисторов к нижнему контакту мемристора. Затем методом магнетронного напыления наносят мемристивный слой, который состоит из оксида переходного метала толщиной в несколько десятков нанометров, и кросс-проводники, соединяющие мемристорные ячейки в параллельную цепь (этап *D* и *E*).

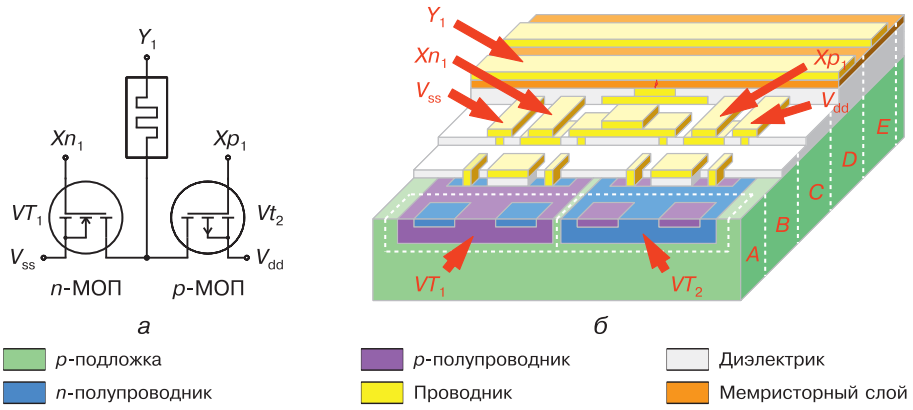



**Рис. 4.1.** Схематическое изображение (*а*) и топология (*б*) мемристорной ячейки с КМОП-транзисторами для блока логического коммутатора

Таким образом, в достаточно простом технологическом процессе формируется сетка мемристорных ячеек, которые через слой диэлектрика и сетки переходных колодцев подключаются к транзисторам, наносимые по стандартной технологии КМОП. Основным преимуществом разработанной матрицы ячеек по сравнению с другими конструкциями энергонезависимых интегральных переключателей является меньший размер ячейки, что позволяет добиться большей степени их интеграции на кристалле.

Важным аспектом проектирования является площадь, занимаемая обвязкой на подложке, которую необходимо разумно расходовать при разработке электрической схемы и проектировании мемристорных ячеек, иначе снижение степени интеграции сведет на нет все преимущества применения мемристоров. Поскольку обойтись в логических схемах без применения транзисторов не представляется возможным, на наш взгляд, оптимальная топология интегральной микросхемы с мемристорами и транзисторами может быть выполнена слоями.

Чтобы обеспечить максимальную степень интеграции в слоистой топологии, занимаемая площадь мемристоров, реализованных в одном слое, должна быть сравнима с занимаемой площадью интегральных транзисторов в другом слое. Исходя из современных технологических возможностей занимаемой площади указанными элементами, можно сделать вывод, что оптимальная конфигурация достигается, когда на один транзистор интегральной схемы приходится около 5—10 мемристоров [14].

Таким образом, соотношение между мемристорами и транзисторами в схеме является показателем ее оптимальности для практической реализации. Для рассмотренной схемы и представленной топологии мемристорной ячейки это соотношение составляет примерно 0,4 мемристора на транзистор (~2,5 транзистора на мемристор). Мемристорные ячейки с КМОП-транзисторами (см. рис. 4.1) объединены в равномерную прямоугольную матрицу. Затворы транзисторов являются входами, предназначенными для ввода переменных в логическую матрицу. Второй контакт мемристора подключен к проводящей линии, которая гальванически объединяет контакты мемристорных ячеек на одной строке для реализации функций конъюнкций и дизъюнкций.

Электрическая схема планарной логической матрицы, реализующей нормальные дизъюнктивные формы, показана на рис. 4.2. Нормальные дизъюнктивные формы позволяют каноническим образом реализовывать любые логические операции, которые требуются для нейроморфного процессора. Количество мемристорных ячеек и подключенных к ним линий растет в соответствии с размерностью матриц, на рисунке показаны только начальные узлы с мемристорными ячейками и линии входов и программирования. Логические переменные подаются на вертикальные линии $X_1$, $\overline{X}_1$, подключенные к затворам входных транзисторов $T_3$, $T_4$, $T_{11}$, $T_{12}$ (см. рис. 4.1). Входные транзисторы образуют КМОП-структуру и с каждым мемристором ($M_1$ и $M_2$) формируют элементарные ячейки, которые выполняют функции коммутации. Мемристоры ячеек, запрограммированные заранее в проводящее состояние, подключают соответствующие логические входы к горизонтальным цепям, а мемристоры в непроводящем состоянии отключают входы от этих цепей. При этом с входными переменными, подключаемыми на одну горизонтальную проводящую линию, выполняется функция конъюнкции.

Полученные литералы на горизонтальных цепях далее подаются на матрицу дизъюнкции, также выполненной на ячейках мемристоров $M_2$, $M_4$ с КМОП-структурами, реализованными на транзисторах $T_5$, $T_6$, $T_{13}$, $T_{14}$. Результатом работы схемы являются выходные логические уровни, полученные как функции последовательной конъюнкции и дизъюнкции от входных переменных по скоммутированным связям мемристорными ячейками.

Программирование мемристоров, составляющих ячейки конъюнктивной и дизъюнктивной матриц, осуществляется с помощью КМОП-транзисторов $T_1$, $T_2$, $T_7$, $T_8$, $T_9$, $T_{10}$, $T_{16}$, $T_{15}$, подключенных к горизонтальным цепям.

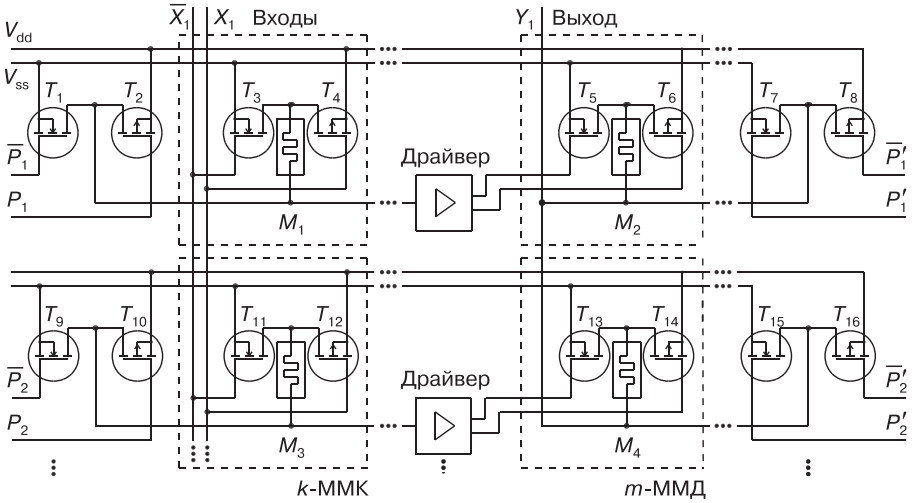



**Рис. 4.2.** Электрическая схема, демонстрирующая способ разделения цепей записи и считывания в логической матрице мемристоров с КМОП-транзисторами:

$k$-ММК — $k$-мерная матрица конъюнкций; $m$-ММД — $m$-мерная матрица дизъюнкций;

$P_1, \overline{P}_1, P_2, \overline{P}_2, P'_1, \overline{P}'_1, P'_2, \overline{P}'_2$ — входы программирования;

$V_{dd}$, $V_{ss}$ — высокий и низкий уровни напряжения питания

Предварительно матрица находится в режиме коммутации логических схем, и оба программирующих транзистора закрыты. Управляющее напряжение, подаваемое на затворы программирующих КМОП-транзисторов, включает режим программирования мемристоров. При этом открывается только один программирующий транзистор верхнего или нижнего плеча в зависимости от того, как требуется перепрограммировать мемристор, ввести его в высокопроводящее или низкопроводящее состояние. Соответственно, для программирования на другом контакте мемристора устанавливается противофазный уровень напряжения через входные КМОП-транзисторы, а их затворами управляют входные логические линии.

Таким образом, представленная схема реализует функцию дизъюнктивной нормальной формы (ДНФ) с возможностью перепрограммируемой коммутации, которая удобна для построения логики работы нейроморфного устройства. Стоит отметить, что недостатком подхода реализации только ДНФ в коммутируемой логике на мемристорах может являться нерациональное использование логических ресурсов матрицы при построении логических схем. С другой стороны, для нейропроцессора этот недостаток может оказаться не столь решающим, потому что требуемые ресурсы будут определяться архитектурой нейронной сети.

Получен патент на электрическую схему и топологию планарной двухслойной логической матрицы на основе мемристорной коммутационной ячейки [15].

Исходя из большого числа нейронов в нейросети и, соответственно, большого числа выходов запоминающей матрицы, размерность логической матрицы должна быть сверхбольшой. Сверхбольшую логическую матрицу в планарной геометрии целесообразно строить из одинаковых модулей с промежуточными усилителями. Количество ячеек в модуле будет ограничено затуханием сигнала из-за паразитных токов. Максимально допустимая деградация логического сигнала будет определяться передаточной характеристикой применяемых усилителей.

## 4.2. ЭЛЕКТРИЧЕСКАЯ СХЕМА, ТОПОЛОГИЯ И НАНОТЕХНОЛОГИЯ ИЗГОТОВЛЕНИЯ 3D ЛОГИЧЕСКОЙ МАТРИЦЫ

Объединение планарных двухслойных (КМОП + мемристорный кроссбар) логических матриц в 3D-структуру обеспечивает высокую интеграцию элементов за счет того, что элементы ячейки и сами ячейки расположены компактно друг над другом [16]. Применение трехмерной организации электрической схемы матрицы значительно уменьшает длину соединительных шин (*buses*) по вертикальным и горизонтальным направлениям, что увеличивает быстродействие и снижает энергопотребление по сравнению с планарной геометрией.

### 4.2.1. Логическая ячейка на основе комбинированного мемристорно-диодного кроссбара

Электрическая схема элементарной ячейки логического матрицы, показанная на рис. 4.3, представляет собой объединение мемристоров с селективными диодами Зенера, подключенные к одному из проводников кроссбара [17]. В свою очередь этот проводник соединен с затвором КМОП-инвертора. Ячейка имеет несколько входов $I_1$–$I_3$, подключенных к затворам полевых транзисторов $T_1$ и $T_2$ через диоды Зенера $D_1$–$D_3$ и мемристоры $M_1$–$M_3$.

Мемристоры подключены к соединенным затворам полевых транзисторов $T_1$ и $T_2$, включенных комплементарно по схеме КМОП-инвертора. Вход инвертора соединен с проводником $P_1$, выходящим на периферию матрицы, при этом каждая ячейка имеет свой, не подключенный к другим ячейкам, проводник, который является цепью программирования мемристоров. Рядом расположенные элементарные ячейки одного уровня соединены с шинами питания $V_{dd}$ и $V_{ss}$, через которые осуществляется управление режимами работы блока. Напряжением питания управляют драйвера, подключенные к входным шинам, которые вынесены на периферию матрицы.

На рис. 4.3 троеточием показано, что мемристорных входов может быть много. Конкретное количество мемристоров зависит от электрических

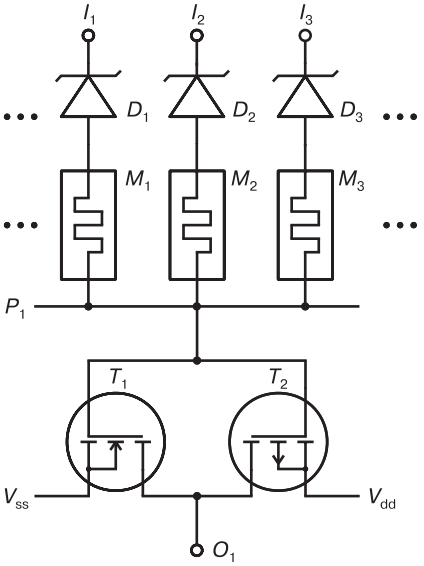



**Рис. 4.3.** Электрическая схема элементарной ячейки для многослойной логической матрицы

свойств и размеров элементов ячейки. Мемристоры следующих ячеек, подключенных на выход инвертора $O_1$, могут находиться в проводящем состоянии в соответствии с его нагрузочной способностью. В основном режиме работы элементарная ячейка питается по линиям $V_{dd}$ и $V_{ss}$ напряжением ниже напряжений туннельного пробоя диодов Зенера и порога переключения мемристора. При низком напряжении питания исключены изменения сопротивлений мемристоров в логической матрице, и мемристоры находятся в режиме хранения своих сопротивлений. В режиме программирования подается напряжение питания на шины $V_{dd}$ и $V_{ss}$ больше напряжений туннельного пробоя диодов Зенера и порога переключения мемристоров, при этом осуществляется программирование матрицы.

Комбинированный кроссбар, состоящий из мемристоров с диодами Зенера, предлагается изготавливать с помощью вакуумной магнетронной технологии. Слои полупроводников с донорной или акцепторной примесью и разным уровнем легирования создаются путем одновременного распыления двух катодов из материалов чистого полупроводника и легирующей примеси [18].

Сначала на подложку через маску наносятся нижние проводники кроссбара $P_1$–$P_8$, а также наращиваются $V_{ss}$, $V_{dd}$ путем распыления металлического катода. Затем пустое пространство между проводниками методом реактивного магнетронного распыления заполняется изолятором (например, диоксидом кремния). Этот слой изолирует питающие шины. Затем реактивным магнетронным распылением наносится пленка оксида переходного металла (например, $TiO_2$), являющаяся мемристорным слоем. Далее через маску последовательно формируются три полупроводниковых слоя диодов Зенера с разной проводимостью в результате одновременного распыления в магнетроне катодов кремния и легирующей примеси.

### 4.2.2. Топология и технология изготовления 3D логической матрицы

Многослойная логическая матрица на основе КМОП-мемристорной коммутационной ячейки представляет собой электронное интегральное устройство на основе логических элементов И–НЕ (NAND), отличающееся

от планарного блока трехмерной архитектурой электрических цепей. На кристалле создается один функциональный пласт, содержащий в нижнем слое КМОП-инверторы, а в верхнем — комбинированный мемристорно-диодный кроссбар. Вышележащий пласт ориентирован ортогонально к нижнему, что является необходимым условием для образования коммутирующих мемристорных кроссбаров между пластами. Такая конфигурация пластов является оптимальной, поскольку позволяет использовать выходные шины пласта в качестве проводников кроссбара.

На рис. 4.4 показан принцип организации топологии 3D логической матрицы [17]. Проводники $I_1$–$I_3$ являются одновременно выходными контактами вышележащего пласта и верхними проводниками кроссбара. Нижние проводники кроссбара $P_1$–$P_6$ присоединены к каждому инвертору по отдельности и выходят на периферию. Выходы инверторов соединены с верхними проводниками кроссбара $O_1$–$O_6$ нижележащего пласта. Мемристорный кроссбар соединяет пласты разного уровня и является верхним слоем ячейки нижнего пласта.

Транзисторы ячейки сформированы в нижележащем слое стандартной КМОП-технологии. Над КМОП-инвертором минимального размера ~$8F^2$ [19] можно разместить до 9 мемристоров. Дальнейшее увеличение числа мемристоров в кроссбаре потребует применения транзисторов большего размера.

Число мемристоров, электрически соединенных с одним инвертором, равно числу синапсов (связей) одного нейрона в запоминающей матрице. Поскольку число синапсов одного нейрона велико (число связей одного нейрона со всеми остальными нейронами сети), то увеличение числа мемристоров, приходящихся на длину одного инвертора, можно обеспечить специальным соединением мемристорного кроссбара [20]. Каждая линия кроссбара при этом будет электрически подсоединена только к одному инвертору (рис. 4.5).

При таком способе соединения достигается высокое соотношение числа мемристоров к одному транзистору, равное 4,5. Для сравнения в массиве Акерса [1] это соотношение равно 0,25. Увеличение общего количества мемристоров на один инвертор за счет включения в схему закрытых мемристоров дает возможность перенаправлять сигнал по многим направлениям и создавать более сложные логические зависимости.

Интеграция мемристорного кроссбара над КМОП-логикой, необходимая для создания одного пласта, уже выполнена в [15]. Последовательное создание пластов 3D сверхбольшой логической матрицы возможно по известной технологии изготовления многослойных чипов, содержащих КМОП-структуры в верхних слоях [21]. Для введения примеси в полупроводниковую подложку при изготовлении транзисторов в нижнем пласте используется быстрое термическое легирование при 1050 °C. КМОП-транзисторы в верхних пластах необходимо изготавливать при температуре 600 °C, чтобы избежать перегрева всей структуры и расплавления в первую очередь

проводников. Для легирования используется метод эпитаксии из твердой фазы. Кроме традиционных этапов при изготовлении транзисторов особым образом формируются проводники шин питания $V_{ss}$, $V_{dd}$ и объединяющий стоки выходной проводник. Сверху на транзисторы инвертора наносится слой изолятора с окнами под вертикальные переходные соединения с нижними проводниками мемристорного кроссбара.

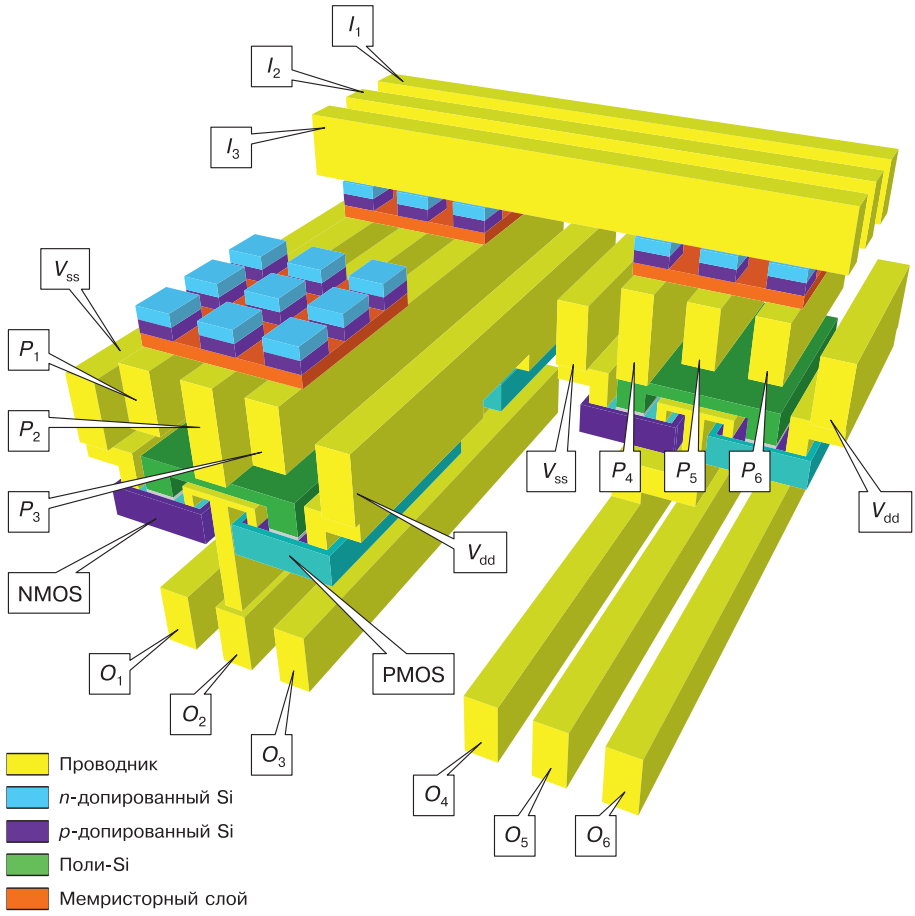



**Рис. 4.4.** Топология 3D логической матрицы с высокой интеграцией элементов

Для сглаживания неровности пластов при их сращивании на нижний пласт необходимо нанести межпластовой диэлектрик, толщиной порядка 100 нм. После изготовления межпластового диэлектрика его верхний слой выравнивается известным методом химико-механической планаризации (Chemical Mechanical Planarization — CMP) [22]. Применение CMP-метода позволяет достичь среднеквадратичной шероховатости подложки, равной 0,2 нм, что намного лучше требований для высококачественного склеивания.

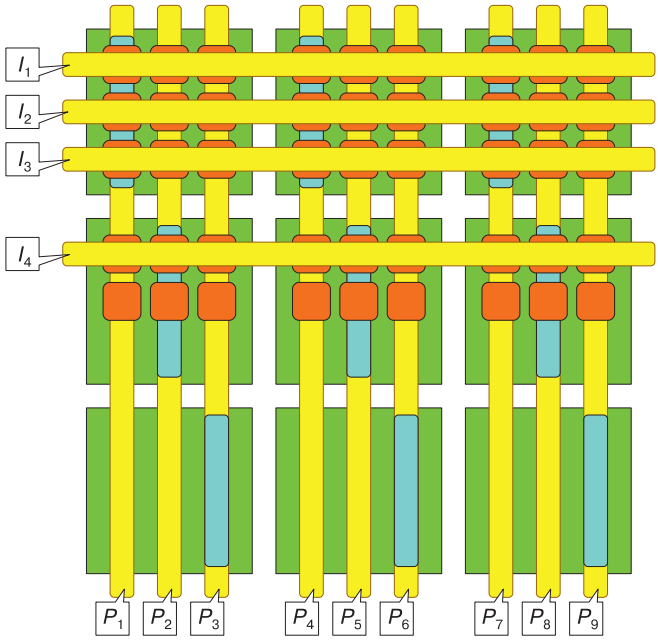



**Рис. 4.5.** Принцип соединения отдельных строк кроссбара с помощью шин с разными инверторами:
зеленый цвет — инверторы; голубой — соединение шины с инвертором

Полированная верхняя поверхность пласта сращивается по технологии низкотемпературного (200 °C) молекулярного склеивания, разработанного для КНИ-подложек [21]. Верхняя подложка гидрофильно приклеивается при комнатной температуре к межпластовому диэлектрику и затем производится управляемое скалывание верхнего слоя кремния по линии предварительно имплантированного водорода. После полировки поверхности скола остается тонкий слой монокристаллического кремния, который служит основой для транзисторов следующего пласта. В полученном слое монокристаллического кремния по технологии межкремниевого соединения TSV (Through-Silicon Via) [23] вытравливаются сквозные колодцы, которые заполняются металлом для верхних проводников мемристорного кроссбара. Эти проводники являются выходами инверторов верхних ячеек и объединяют катоды диодов нижележащей ячейки, причем выходы инверторов разных ячеек одного пласта не соединяются между собой.

### 4.2.3. Электрическая схема матрицы

Многослойная логическая матрица, состоящая из КМОП-инверторов и мемристоров, в основном режиме работы реализует булеву функцию, построенную на основе последовательного выполнения логических операций, на выходной линии и в инверторе. Электрическая схема функционального

пласта из трех ячеек на основе инверторов $DD_4$–$DD_6$ показана на рис. 4.6 [17]. Инверторы $DD_1$–$DD_3$ являются частью другого пласта.

Для программирования кроссбара текущего пласта необходимо отключить питание всех инверторов, кроме входных по отношению к текущему пласту. Программирование мемристоров $M_1$–$M_9$ осуществляется по шинам $P_1$–$P_3$, являющимся проводниками мемристорного кроссбара и подключенными к входам соответствующих инверторов $DD_4$–$DD_6$. К противоположным контактам мемристоров напряжение прикладывается от инверторов $DD_1$–$DD_3$. Независимое управление входными инверторами позволяет применить обычный алгоритм [24] записи в кроссбар с пассивными селекторами.

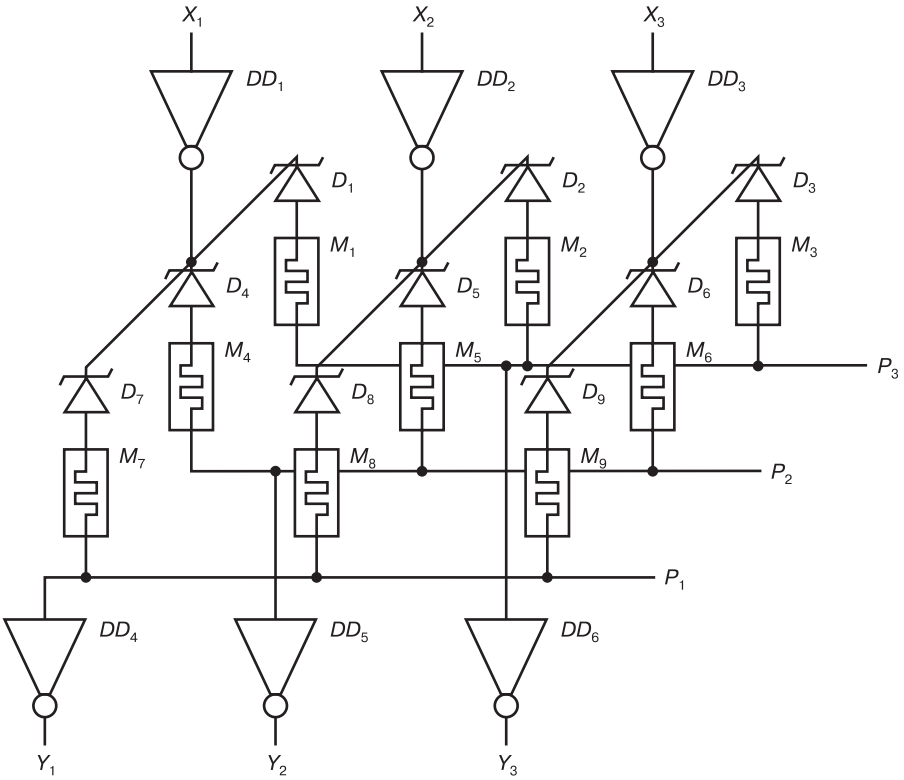



**Рис. 4.6.** Электрическая схема функционального пласта с тремя ячейками и межпластовое соединение ячеек в логическом блоке

В отличие от материальной импликации [4–7; 10] в представленном логической матрице изменение состояния мемристоров происходит только во время программирования логических функций, что увеличивает срок службы устройства. Использование активных элементов в ячейке устраняет проблему деградации сигнала при выполнении множества операций в комбинационной схеме, характерную для полностью пассивной схемы [1].

Получен патент на электрическую схему и топологию 3D сверхбольшой логической матрицы на основе комбинированного мемристорно-диодного кроссбара [25].

## 4.3. МАРШРУТИЗАЦИЯ ВЫХОДНЫХ СИГНАЛОВ НЕЙРОННОГО БЛОКА

Рассмотрим работу логической матрицы в режиме маршрутизации сигналов. Нейроны запоминающего блока нейропроцессора посылают в логический блок одинаковые по форме импульсы, следовательно, для их маршрутизации можно использовать цифровую схемотехнику. Поэтому логический блок может работать в качестве маршрутизатора, направляя выходные импульсы нейронов на синапсы других нейронов. Причем инверторы усиливают сигнал, т. е. коэффициент объединения по выходу у нейрона будет большой, что полезно при сверхбольшом размере блока.

Схему, выполняющую маршрутизацию сигналов, можно реализовать с помощью логической матрицы из двух функциональных пластов. Первый пласт логической матрицы запрограммирован таким образом, что низкое сопротивление имеют мемристоры на главной диагонали. Следовательно, в первом пласте образован набор одновходовых элементов И–НЕ, эквивалентных элементу НЕ. Мемристоры с низким сопротивлением во втором пласте логической матрицы подключают требуемые выходы из первого пласта к входной шине выходного инвертора. Один инвертор с подключенными к нему мемристорами выполняет конъюнкцию с отрицанием над инвертированными в первом пласте сигналами, что в итоге эквивалентно дизъюнкции над ними.

В результате логическая матрица из двух функциональных пластов позволяет перенаправлять на произвольный его выход сигналы с любых его входов. Маршруты для сигналов, запрограммированные в логическом блоке, однозначно определяют архитектуру нейросети, которая будет установлена на нейропроцессор.

## 4.4. УМНОЖЕНИЕ МАТРИЦЫ ЧИСЕЛ НА ВЕКТОР С ИСПОЛЬЗОВАНИЕМ ПОЗИЦИОННОГО КОДИРОВАНИЯ

Рассмотрим работу логической матрицы [17] в режиме умножения матрицы чисел на вектор. В предлагаемой схеме, в отличие от аналоговой схемы умножения матрицы на вектор с применением промежуточных состояний мемристоров, умножение построено с использованием режима маршрутизации при позиционном импульсном (цифровом) кодировании сигнала. Такое кодирование означает, что с каждым значением компоненты входного

вектора связана уникальная входная электрическая шина в логической матрице. Ввод сигнала в схему осуществляется путем подачи коротких импульсов на соответствующие шины (рис. 4.7).

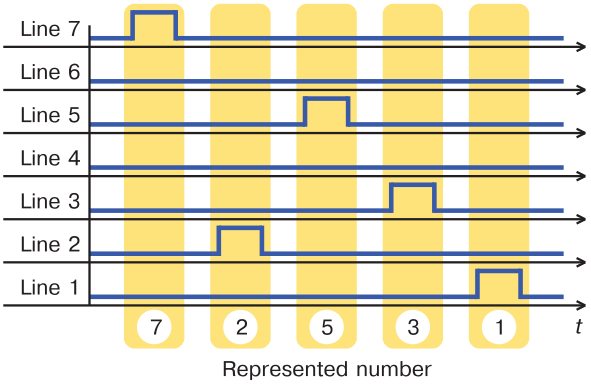



**Рис. 4.7.** Принцип позиционного кодирования чисел

Количество одновременно подаваемых импульсов и, соответственно, активных входных шин в данном случае равно числу компонент входного вектора. Задача умножения сводится к правильной маршрутизации входных импульсов на соответствующие результату умножения выходные шины.

Импульсы, проходя через слои 3D логической матрицы, распространяются по мемристорным кроссбарам и в зависимости от заранее запрограммированной проводимости мемристоров появляются на определенных выходных инверторах.

Умножение строки матрицы на вектор-столбец разбивается на две операции: умножение элемента матрицы на компонент вектора и суммирование получившихся произведений. Первая операция выполняется в двух функциональных пластах логического блока путем перенаправления импульса с входной шины на выходную шину, отвечающую произведению элемента матрицы на компоненту входного вектора. Вторая операция выполняется путем организации в функциональном пласте логического блока элементов И–НЕ, соответствующих всем возможным комбинациям слагаемых. В следующем функциональном пласте происходит перенаправление импульсов от выходов предыдущего пласта, соответствующих одинаковым числам, на одну выходную шину с использованием операции ИЛИ–НЕ.

В итоге входные импульсы проходят на определенные выходные шины, которые соответствуют результату операции матрично-векторного умножения. Запись значений числовой матрицы в логическую матрицу заключается в программировании мемристоров функциональных пластов в низкопроводящее и высокопроводящее состояния, так чтобы выходные значения соответствовали результату умножения.

В многослойной КМОП-мемристорной логической матрице в случае позиционного импульсного кодирования входной информации реализуется рекуррентная формула

$$\vec{Y} = M_n\left(\ldots\left(M_2\left(M_1\vec{X}\right)\right)\ldots\right).$$

Следует отметить, что рекуррентная векторная формула имеет значительный объем математических вычислений. В частности, это приводит к низкому быстродействию и высокому энергопотреблению в случае расчета многослойной нейросети на микропроцессоре с традиционными архитектурами. Представленная 3D логическая матрица лишена этих недостатков. Она позволяет производить вычисления рекуррентной формулы параллельно и с высоким быстродействием. При этом работа транзисторов матрицы происходит в ключевом режиме, который характеризуется малым энергопотреблением.

## 4.5. ЛОГИЧЕСКАЯ МАТРИЦА ВО ВХОДНОМ/ВЫХОДНОМ БЛОКЕ НЕЙРОПРОЦЕССОРА

Входной блок предназначен для первичной обработки звуковых и видео сигналов, включающей в себя фильтрацию и кодирование информации в импульсы, а также для кодировки любой другой информации в виде отдельных импульсов и, если потребуется, для преобразования этих импульсов в импульсы с определенными амплитудой и длительностью, подобные биоморфным импульсам мозга. Под фильтрацией подразумевается дискретное косинусное преобразование и отсев малых амплитуд.

По окончании процессинга в запоминающей и логической матрицах, информация поступает на выходной блок, где происходит ее окончательная обработка (спектральный анализ, сжатие изображения и сверточная фильтрация). Далее информация, подготовленная к транспортировке, поступает на интерфейсный блок.

В большинстве алгоритмов сжатия и распаковки аудиовизуальной информации используется процедура умножения матрицы чисел на вектор, реализованная в предлагаемой логической матрице. Кроме того, умножение матрицы чисел на вектор является частью процедуры дискретного косинусного преобразования, используемого, например, в алгоритме сжатия и хранения изображений в формате JPEG на компьютере. Это преобразование является разновидностью преобразования Фурье и заключается в скалярном умножении входного вектора (например, пикселей картинки) на числовую матрицу преобразования. Умножение может быть выполнено в цифровом или в аналоговом видах. Цифровой подход обеспечивает большую точность, но требует значительно большего числа элементов. Наиболее продвинутый массив мемристорных ячеек, впервые предложенный

Hewlett-Packard как аппаратное средство для выполнения дискретного косинусного преобразования при обработке сигналов (сжатие изображений и сверточная фильтрация) [12]. Аналоговое умножение в этой матрице основано на использовании закона Ома для умножения входного напряжения на проводимость мемристора, отвечающей значению элемента матрицы преобразования, и последующего сложения получившихся токов на общей выходной шине по закону Кирхгофа. В электрической схеме массива [12] последовательно с каждым мемристором включен селективный транзистор, используемый только при записи состояния мемристоров. Для реализации фильтрации необходимо выполнить отсев малых амплитуд из образа преобразования. Процедуру умножения числовой матрицы на вектор выполняет аналогичный массив [26].

Умножение матрицы на вектор в цифровом виде возможно с применением логической матрицы [17], построенной на основе мемристорно-диодного кроссбара с ячейками 1D1M, в которых в качестве селективного элемента использован диод Зенера.

## 4.6. ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ РАБОТЫ ЛОГИЧЕСКОЙ МАТРИЦЫ

При моделировании работы логичекой матрицы использовалась оригинальная специализированная программа MDC-SPICE, разработанная для расчета больших электрических схем с мемристорно-диодными кроссбарами.

Для определения максимально возможного размера кроссбара логической матрицы было проведено моделирование процессов деградации логических уровней, вызванных паразитными токами в комбинированном мемристорно-диодном кроссбаре. Входы инверторов $P_1$, $P_2$, ... (см. рис. 4.6) при этом были подтянуты к положительному полюсу питания резисторами 200 кОм. Максимально возможная деградация определяется передаточной характеристикой инверторов. Для инвертора, образованного из приведенных моделей транзисторов, порог равен 65 %.

Поскольку выходы функционального пласта логического блока идентичны, то достаточно рассмотреть деградацию логических уровней (логические 0 и 1) на входе одного инвертора. На рис. 4.8 красным цветом показана деградация напряжения логического нуля в разработанной матрице с ячейками 1D1M, а синим и зеленым — деградация напряжения при двух соотношениях сопротивлении открытого и закрытого мемристоров ($R_{off}/R_{on}$ = 100 и 1000) в логическом массиве ячеек 4T1M Levy [1], который имеет выход на один инвертор. Красная кривая достигает 80 % уровня первоначального напряжения в 1 В из-за влияния диода Зенера, открывающегося в прямом смещении при 0,3 В.

Из рис. 4.8 видно, что с ростом размера кроссбара предлагаемого блока напряжение выходного сигнала медленно уменьшается от уровня 80 % по линейному закону и достигает предельных 65 % при размере $N \times N = 10^{10}$.

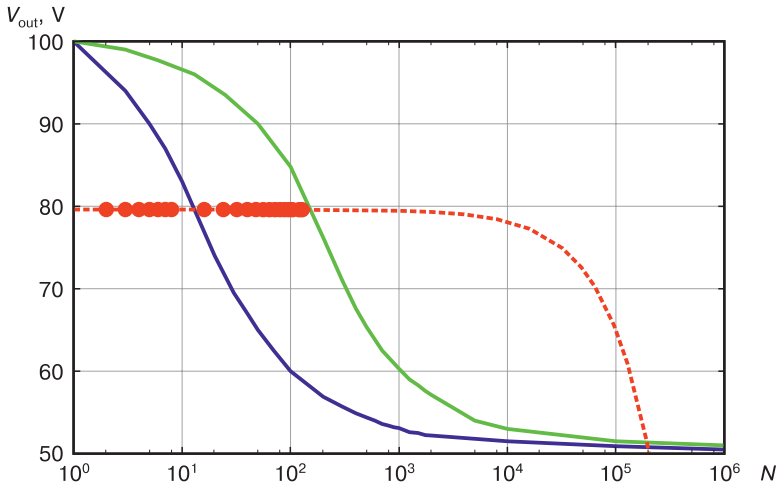



**Рис. 4.8.** Затухание логического сигнала от количества ячеек в массивах размером $N \times N$:

красная кривая ($R$ = 100 и 1000) — для разработанного блока; синяя ($R$ = 100) и зеленая ($R$ = 1000) кривые — данные из [1]

А максимальный размер массива Levy [1] 500 × 500 достигается при $R = 1000$.

Для тестирования работоспособности 3D логической матрицы было проведено SPICE-моделирование фрагмента, состоящего из двух пластов, которые содержат по две логические ячейки [16]. Каждая ячейка фрагмента содержит два коммутирующих мемристора. Информация в мемристоры записывалась последовательно. Высокое сопротивление получили мемристоры $M_1$ и $M_5$, а остальные — низкое. В результате на верхнем слое реализовались логические функции $Y_{11} = V(x_{111}) = X_1 \vee X_2$ и $Y_{12} = V(x_{112}) = X_2$, а на нижнем слое $Y_{21} = \text{НЕ}(Y_{11}) = \text{НЕ}(X_1 \vee X_2)$ и $Y_{22} = \text{НЕ}(Y_{12}) = \text{НЕ}(X_2)$.

На рис. 4.9 показаны входные и выходные уровни напряжения матрицы, полученные в ходе моделирования.

Из рис. 4.9 следует, что выходные сигналы матрицы, полученные в ходе моделирования, соответствуют запрограммированным функциям: сигнал $Y_{11}$ отражает дизъюнкцию $X_1$ и $X_2$, а $Y_{12}$ — инверсию $X_2$. Выходные сигналы $Y_{21}$, $Y_{22}$ стробированы импульсами, приходящими на питание выходных инверторов.

Для проверки работоспособности режима умножения матрицы на вектор было проведено моделирование электрической схемы, выполняющей умножение матрицы размером 3 × 3 на трехкомпонентный вектор [17]. Реализация умножения в нескольких функциональных пластах, реализующих конъюнкцию с инверсией, возможна при использовании позиционного кодирования чисел. Каждый вход и выход схемы отвечает за конкретное числовое значение.

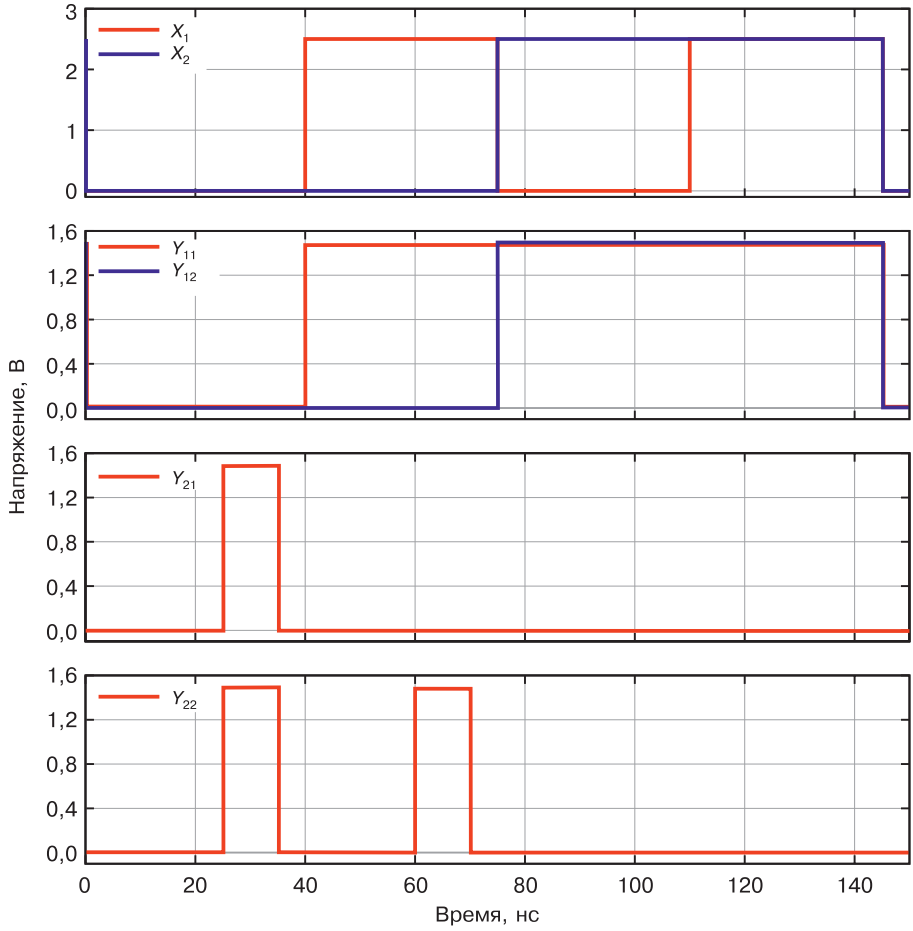



**Рис. 4.9.** Эпюры напряжения на входе и выходе пластов логической матрицы:

*а* — вход верхнего пласта; *б* — выходы верхнего пласта/входы нижнего пласта; *в* — выход нижнего пласта; *г* — выход нижнего пласта

Схема состоит из четырех функциональных пластов, обозначенных на рис. 4.10 разными цветами. Синий блок выполняет инверсию входных сигналов, светло-зеленый — непосредственно умножение компонента вектора на элемент матрицы, путем перенаправления импульса на соответствующую шину. Блок суммирования состоит из двух частей: темно-зеленые блоки представляют собой набор трехвходовых элементов И–НЕ, соответствующих уникальным комбинациям (суммам) полученных произведений, а сиреневые блоки передают на выход уникальные суммы.

В эксперименте возможные значения компонент вектора и элементов матрицы представлены целыми числами и принадлежат диапазону от 1 до 3. Уровень напряжения логической единицы равен 1,5 В, логического

нуля — 0 В. Вход каждого инвертора подсоединен к источнику питания через подтягивающий резистор 200 кОм. На рис. 4.10 представлен результат моделирования умножения трех векторов: (1, 2, 3), (2, 1, 3) и (3, 2, 2) на матрицу *M*, которая запрограммирована в виде сопротивлений мемристоров во втором функциональном слое. Сопротивление мемристоров в высокопроводящем и низкопроводящем состояниях равны соответственно 1 кОм и 10 МОм Умножение трех векторов производится последовательно во времени. На рис. 4.10 не показано изменение напряжений на всех 66 выходах, поскольку импульсы возникают только на линиях, соответствующих компонентам выходного вектора.

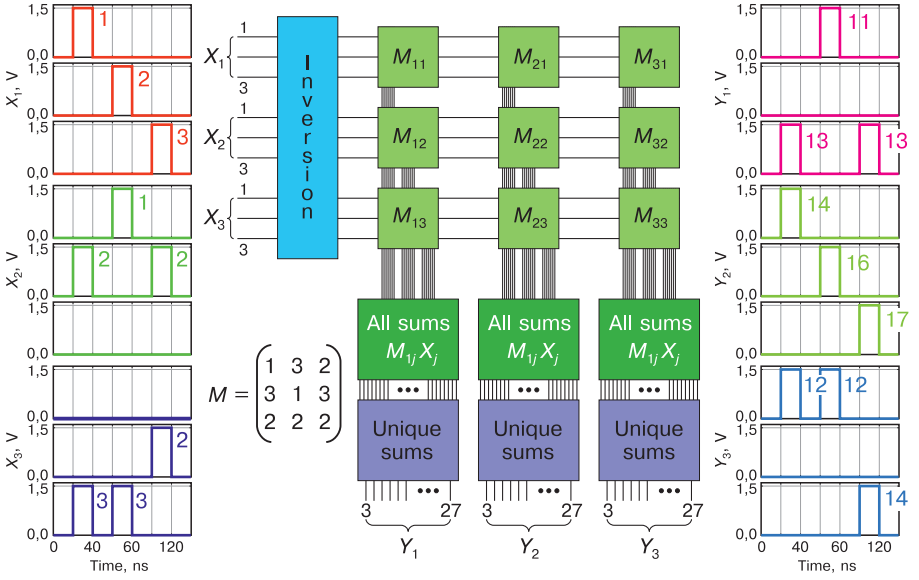



**Рис. 4.10.** Моделирование умножения матрицы чисел на вектор в логической матрице

Как видно из диаграмм, показанных справа на рис. 4.10, выходной вектор соответствует ожидаемому результату, вычисленному по правилу матрично-векторного умножения. Электрическая схема является комбинационной и выполняет умножение за один такт.

Моделирование режима маршрутизации сигналов производилось для логической матрицы (рис. 4.11), состоящей из двух функциональных пластов с комбинированным кроссбаром размером 3 × 3 [27]. На входах каждого инвертора был установлен подтягивающий к напряжению питания резистор с сопротивлением 100 кОм. На три первых входа были поданы уникальные последовательности импульсов: для первого входа — один импульс, для второго — группа из 2, а для третьего — группа из 3. Диагональные мемристоры первого функционального пласта запрограммированы

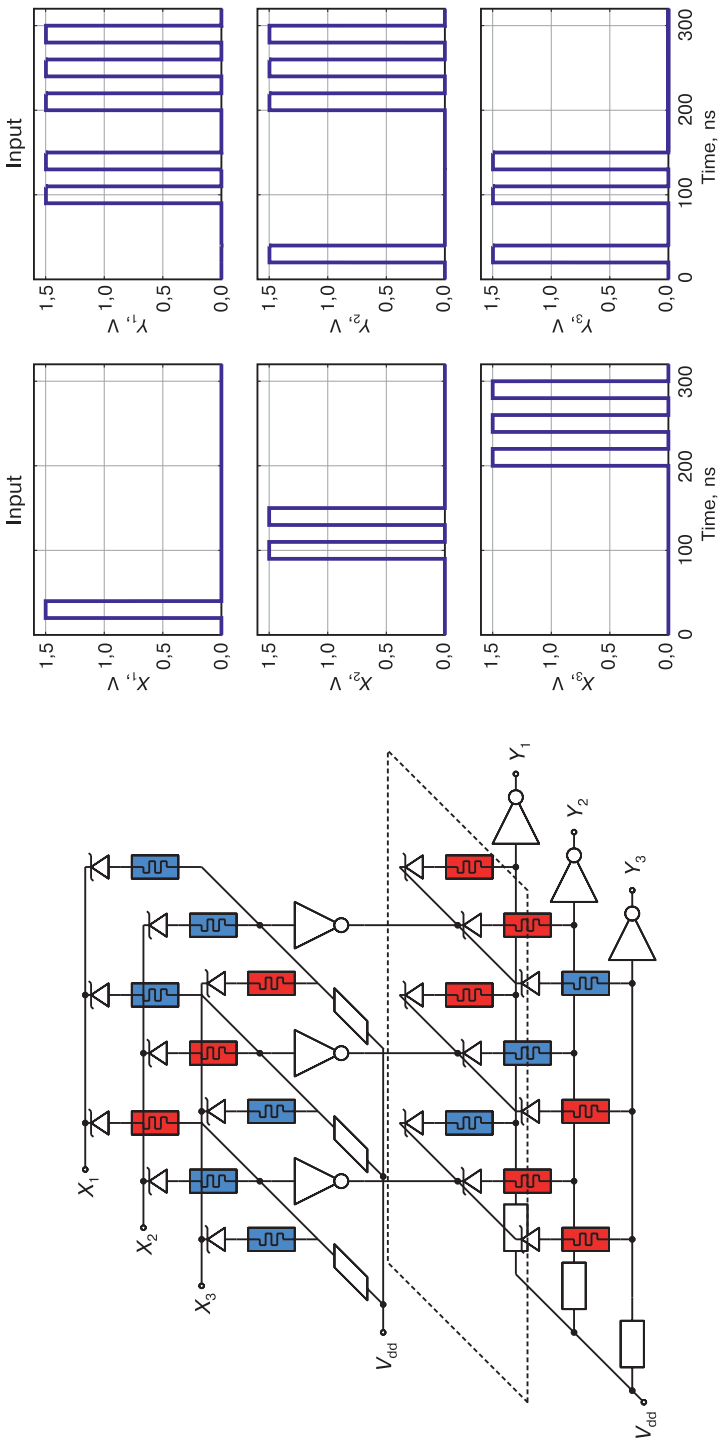



**Рис. 4.11.** Маршрутизация сигналов в логической матрице, состоящей из двух пластов

в высокопроводящее состояние (10 кОм) для образования одновходовых элементов И–НЕ и инверсии входных импульсов. Сопротивление низкопроводящих мемристоров при этом было равно 1 МОм.

Второй функциональный пласт запрограммирован следующим образом: мемристоры с низким сопротивлением на каждой выходной линии образуют элементы 2И–НЕ, принимающие на вход инвертированные в первом пласте импульсы, причем на первый выход направляются второй и третий входы, на второй выход — первый и третий входы, на третий выход — первый и второй входы. Выходные напряжения, представленные на рис. 4.10 показывают успешное перенаправление входных сигналов по заданным маршрутам.

Последнее моделирование было направлено на сравнение энергопотребления массива Hewlett-Packard (HP) [12] размером 2 × 2 при умножении матрицы размером 2 × 2 на двухкомпонентный вектор и равнофункционального разработанного логического блока, состоящего из одного функционального пласта размером 12 × 12 [28]. Для простоты во втором случае элементы матрицы и компоненты вектора были представлены двоичными 6-битными числами. Умножение выполнено путем сдвига компонент входного вектора логического блока в режиме маршрутизации сигналов, что эквивалентно умножению на степень двойки.

На рис. 4.12 показана временная зависимость потребляемой энергии при обработке входного сигнала диодно-мемристорной матрицей размером 24 × 24 ячеек в позиционном коде и 16 ячейками аналоговой матрицы HP [12] с возможным числом состояний, равным 64, что эквивалентно 6 битам.

Моделирование массива HP выполнялось с использованием модели операционных усилителей AD8031 с двухполярным питанием ±3 В. Сопротивление в преобразователе ток-напряжение равно 4 кОм. Амплитуды компонент вектора равны 16, 32, 64, 128 мВ. Сопротивление закрытых мемристоров составляет 100 кОм, открытых — 1 кОм. В цифровой логической матрице подтягивающее к питанию сопротивление равно 100 кОм. Характеристики мемристоров те же.

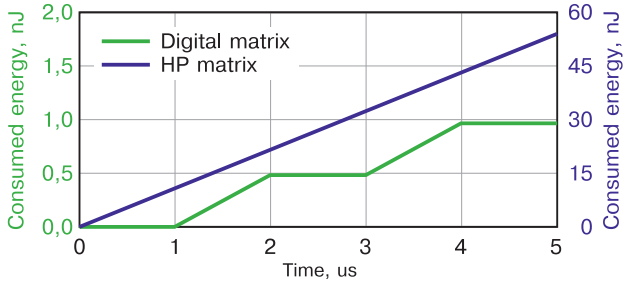



**Рис. 4.12.** Изменение потребленной энергии во времени:
синяя кривая — аналоговая матрица HP;
зеленая кривая — предлагаемая цифровая логическая матрица

Из рис. 4.12 следует, что общий объем потребленной энергии матрицей НР непрерывно возрастает во времени, в то время как в предлагаемой матрице рост происходит преимущественно в моменты переключений инверторов. Так, после прохождения второго импульса величина энергии, потребленной матрицей НР, превышает соответствующую величину цифровой матрицы в 56 раз. Основными потребителями энергии в матрице НР являются операционные усилители (2,7 мВт на каждый ОУ), а в разработанной матрице — КМОП-инверторы. Поскольку аналоговая матрица в процессе работы постоянно потребляет энергию, а цифровая — только в моменты изменения напряжения на входах, то с увеличением числа ячеек разница в энергопотреблении НР и диодно-мемристорной матриц будет возрастать.

## 4.7. РЕАЛИЗАЦИЯ НЕЙРОННЫХ ФУНКЦИЙ ЗАПОМИНАЮЩЕЙ МАТРИЦЫ В ЛОГИЧЕСКОЙ МАТРИЦЕ

Нейронные функции сложения и вычитания взвешенных сигналов, которые реализует запоминающая матрица, могут быть выполнены в логической матрице при отсутствии самой запоминающей матрицы, согласно варианту II нейропроцессора (см. рис. 1.18). Для этого аналогично развитой электрической модели нейрона в запоминающей матрице (см. рис. 1.19) необходимо провести усовершенствование биоморфной электрической модели нейрона в логической матрице [27].

На периферии логической матрицы установлена электрическая цепь обратной связи с транзисторами, выполняющими функции управляющих драйверов [28]. В этой цепи к шинам, соединенным с затворами КМОП-транзисторов, дополнительно подключены периферийные конденсаторы. Эти конденсаторы при необходимости могут быть подключены к паразитным шинным емкостям и емкостям затворов инвертора, если последних окажется недостаточно. Каждый конденсатор заряжается электрическими импульсами потенциала действия и выполняет функцию клеточной мембраны биологического нейрона.

Импульсы напряжения, проходя сквозь предыдущий инвертирующий пласт, поступают на конденсаторы через единичные мемристоры, проводимость которых является весом для суммируемых сигналов. Образовавшиеся после взвешивания в мемристорах токи складываются на общей выходной шине и заряжают конденсаторы. Порог активации нейронов определяется напряжением срабатывания КМОП-инверторов логической матрицы.

Функциональная схема одного нейрона логической матрицы показана на рис. 4.13. Уменьшение напряжения на суммирующих конденсаторах внутри логической матрицы, моделирующей процесс возбуждения нейрона, происходит за счет входных импульсов. Импульсы низкого логического уровня через соответствующий мемристор и диод Зенера уменьшают

электрическое напряжение на этих конденсаторах и при достижении порога срабатывания инвертора происходит формирование переднего фронта выходного импульса. Формирование заднего фронта выходного импульса осуществляет схема обратной связи на периферии, построенная на RC-цепи и двухпороговом компараторе, который через транзистор заряжает конденсаторы, тем самым сбрасывая возбуждение нейрона.

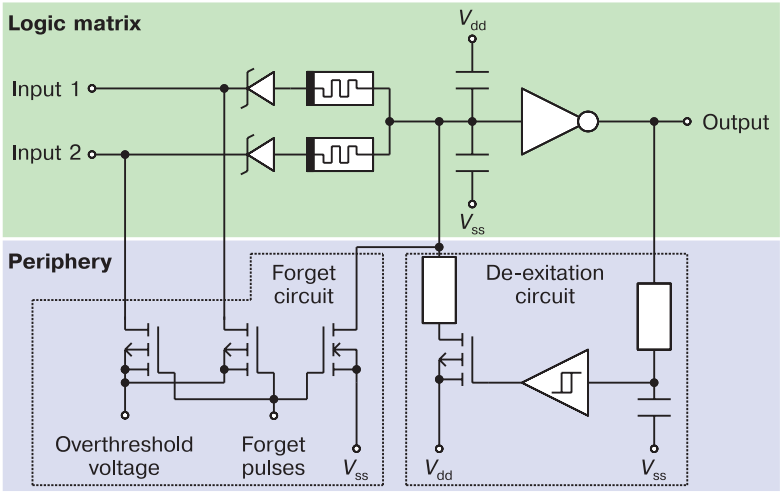



**Рис. 4.13.** Развитая электрическая схема нейрона в логической матрице: зеленый цвет — внутренние элементы матрицы; синий — периферийные элементы

Достоинство этой схемы нейрона по сравнению с аналогичной схемой с использованием запоминающей матрицы (см. рис. 1.19) заключается в меньшем энергопотреблении, поскольку в ней отсутствует работающий в аналоговом режиме источник тока (CCCS). С другой стороны, отсутствие комплементарной пары мемристоров в логической матрице позволяет использовать только положительные синаптические веса. Реализация отрицательных весов, необходимых для торможения нейронов, требует доработки схемы.

## 4.8. АССОЦИАТИВНОЕ САМООБУЧЕНИЕ СИНАПСОВ В ЛОГИЧЕСКОЙ МАТРИЦЕ И ГЕНЕРАЦИЯ НОВОЙ АССОЦИАЦИИ

Ассоциативное самообучение логической матрицы [27; 28]. происходит аналогично обучению запоминающей матрицы. Обучение аппаратной нейронной сети выполняется в соответствии с правилом Хебба путем программирования мемристоров соответствующих связей импульсами надпорогового напряжения. Такие импульсы, приводящие к усилению синаптической

связи, формируются выходом нейрона в процессе условного обучения сети. Для этого к шинам питания инверторов присоединены драйвера, поднимающие входное напряжение выше порога программирования мемристоров. Активация драйверов происходит во время появления выходных импульсов нейрона, что приводит к пробою диода Зенера и уменьшению электрического сопротивления тех мемристоров, на которых присутствуют в это время входные информационные импульсы.

Чтобы логический блок нейропроцессора не насыщался большим количеством связей между нейронами, был реализован процесс безусловного саморазобучения нейронной сети. Это аналогично нейрофизиологическому явлению забывания редко используемой информации. Для реализации процесса безусловного разобучения от внешнего генератора периодически поступают на мемристоры разобучающие импульсы надпорогового напряжения, увеличивающие электрическое сопротивление мемристоров.

На рис. 4.14 представлены результаты MDC–SPICE-моделирования аппаратной нейросети, имеющей слоистую структуру, в которой сигнал распространяется перпендикулярно к поверхности подобно кортикальной колонке мозга. Нейросеть состоит из трех нейронных слоев, реализованных в шести функциональных пластах логической матрицы. Связи между нейронами, отмеченные синим цветом, являются слабыми, а связи, отмеченные красным, — сильными.

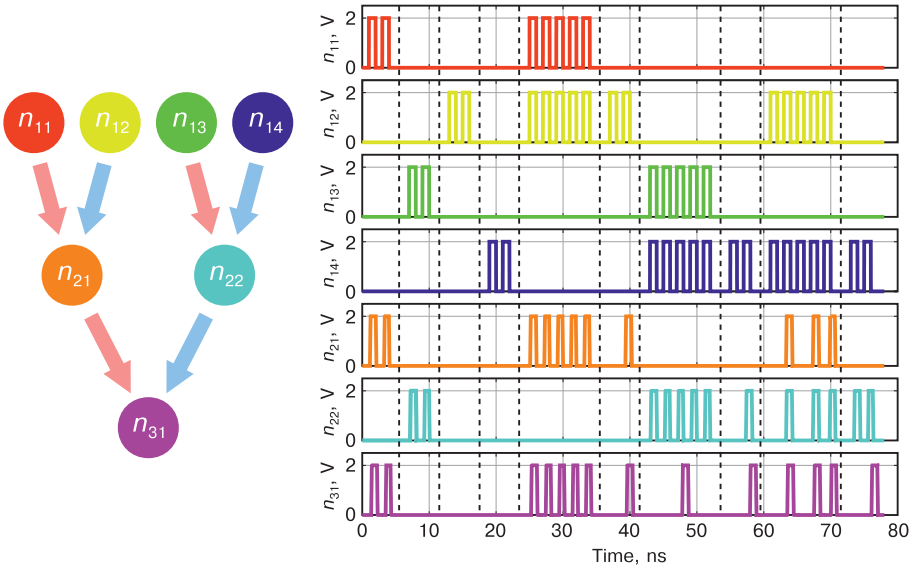



**Рис. 4.14.** Ассоциативное обучение
в трехслойной аппаратной нейросети:
*а* — архитектура трехслойной аппаратной нейросети;
*б* — эпюры SPICE-моделирования
(цвет графика соответствует цвету нейрона)

Временные области V, VII и IX на рис. 4.14 демонстрируют усиление слабых нейронных связей $n_{12}$–$n_{21}$, $n_{14}$–$n_{22}$ и $n_{22}$–$n_{31}$ путем последовательного возникновения новых ассоциаций. В результате, как видно из временной области *X*, происходит ассоциация нейрона $n_{14}$ с нейроном $n_{31}$ через промежуточный слой нейронов. Результат моделирования показывает, что механизм обучения является транзитивным и работает не только между соседними слоями нейронов, а распространяется на всю аппаратную нейросеть. Уменьшение частоты импульсов выходного нейрона $n_{31}$ в конце графика является следствием работы схемы разобучения. В этом режиме происходит безусловное ослабление усиленных ранее связей, которые не подкреплялись ассоциативным самообучением в предыдущей временной интервал.

На рис. 4.15 показаны кривые изменения электрического сопротивления мемристоров, изначально слабых нейронных связей, полученных путем SPICE-моделирования логической матрицы. В моменты самообучения длительностью 80 нс происходит подача на мемристоры импульсов надпорогового напряжения программирования, приводящих к уменьшению их электрического сопротивления. Переход слабой нейронной связи в сильную является пороговым процессом. Этот порог задается параметрами управляющей схемы логической матрицы, располагающейся на периферии.

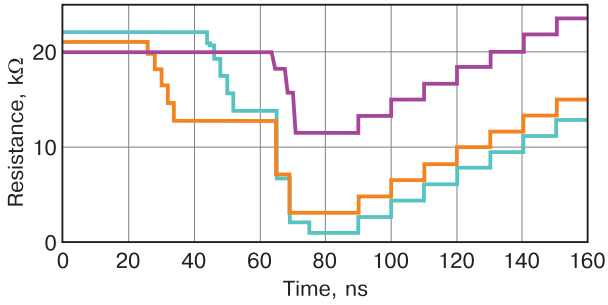



**Рис. 4.15.** График изменения электрического сопротивления мемристоров, имитирующих работу синапсов при ассоциативном самооообучении

Обратный процесс — увеличение электрического сопротивления мемристоров — имитирует процесс безусловного разобучения нейронной сети в логическом блоке. Он показан на рис. 4.15 в интервале времени от 80 до 160 нс. В этом интервале работает часть схемы, отвечающей за разобучение логической матрицы. Каждый импульс генератора разобучения приводит к увеличению электрического сопротивления соответствующего мемристора.

На рис. 4.16 отражно энергопотребление электрической схемы из трех нейронов ($n_{11}$, $n_{12}$ и $n_{21}$) в логической матрице при обучении (см. рис. 4.14). Три нейрона выбраны для сравнения с энергопотреблением трех нейронов в запоминающей матрице (см. рис. 3.12).

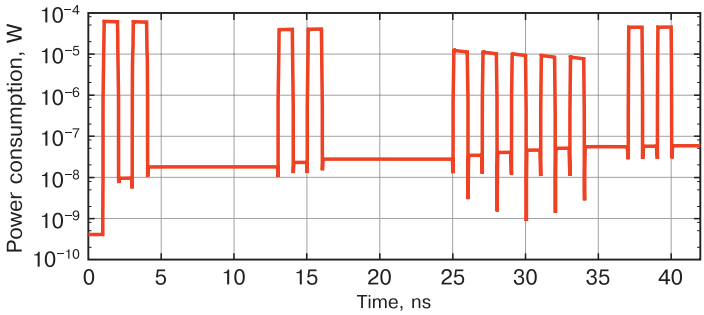



**Рис. 4.16.** Изменение во времени потребляемой мощности схемы из трех нейронов при обучении синапсов логической матрицы

Аналогично результату моделирования запоминающей матрицы, энергопотребление логической матрицы, работающей в режиме запоминающей, резко возрастает при наличии потенциалов действия. Короткие всплески во время передних фронтов соответствуют переключению инверторов (кратковременная работа в активном режиме). Средняя потребляемая электрической схемой из трех нейронов мощность за время моделирования составила 30,3 мкВт, что в 3,5 раза меньше потребления запоминающей матрицы.

Таким образом, пласты логической матрицы, функционально дополненные электрическими цепями на периферии, позволяют реализовать механизмы условного самообучения и безусловного разобучения аппаратной нейронной сети нейропроцессора. Слои нейронов в логической матрице могут быть использованы для имитации работы колонки кортекса, с перпендикулярным к поверхности распространением информационных сигналов подобно биологической нейронной архитектуре.

## Список литературы

1. *Levy Y.*, *Bruck J.*, *Cassuto Y.* et al. Logic operations in memory using a memristive Akers array // Microelectronics Journal. 2014. Vol. 45. Pp. 1429–1437.
2. *Strukov D.B.* Hybrid CMOS/nanodevice circuits with tightly integrated memory and logic functionality // Nanotechnology 2011: Electronics, Devices, Fabrication, MEMS, Fluidics and Computational, CRC Press, Boca Raton, Florida, USA. 2011. Pp. 9–12.
3. *Xia O.*, *Robinett W.*, *Cumbie M.W.* et al. Memristor — CMOS hybrid integrated circuits for reconfigurable logic // Nano Letters. 2009. Vol. 9. Pp. 3640–3645.
4. *Lehtonen E.*, *Laiho M.* Stateful Implication Logic with Memristors // IEEE/ACM International Symposium on Nanoscale Architectures. 2009. Pp. 33–36.
5. *Linn E.*, *Rosezin R.*, *Tappertzhofen S.* et al. Beyond von Neumann-logic operations in passive crossbar arrays alongside memory operations // Nanotechnology. 2012. Vol. 23. P. 305205.
6. *Kvatinsky S.*, *Satat G.*, *Wald N.* et al. Memristor — based material implication (IMPLY) logic: Design principles and methodologies // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2014. Vol. 22. No. 10. Pp. 2054–2066.

7. *Prezioso M.*, *Riminucci A.*, *Graziosi P.* et al. A single-device universal logic gate based on a magnetically enhanced memristor // Advanced Materials. 2013. Vol. 25. Pp. 534–538.
8. *Kvatinsky S.*, *Belousov D.*, *Liman S.* et al. MAGIC — memristor-aided logic // IEEE Transactions on Circuits and Systems — II: Express Briefs. 2014. Vol. 64. No. 11. Pp. 895–899.
9. *Kvatinsky S.*, *Wald N.*, *Satat G.* et al. MRL — memristor ratioed logic // 13th International Workshop on Cellular Nanoscale Networks and their Applications. 2012. P. 13073664.
10. *Adam G.C.*, *Hoskins B.D.*, *Prezioso M.*, *Strukov D.V.* Optimized stateful material implication logic for 3D data manipulation // Nano Research. 2016. Vol. 9. No. 12. Pp. 3914–3923.
11. *Yang J.J.*, *Strukov D.B.*, *Stewart D.R.* Memristor devices for computing // Nature Nanotechnology. 2013. Vol. 8. Pp. 13–24.
12. *Li C.*, *Hu M.*, *Li Y.* et al. Analogue signal and image processing with large memristor crossbars // Nature electronics. 2018. Vol. 1. No. 1. Pp. 52–59.
13. *Маевский О.В.*, *Писарев А.Д.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* Логический коммутатор и запоминающее устройство на основе мемристорных ячеек для электрической схемы нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2016. No. 4. С. 100–111.
14. *Ракитин В.В.* Интегральные схемы на КМОП-транзисторах: учеб. пособие. — М., 2007. 307 с.
15. *Маевский О.В.*, *Писарев А.Д.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* Логическая матрица на основе мемристорной коммутационной ячейки. 2018. Патент No. 2643650.
16. *Udovichenko S.*, *Pisarev A.*, *Busygin A.*, *Maevsky O.* 3D CMOS, memristor nanotechnology for creating logical and memory matrics of neuroprocessor // Nanoindustry. 2017. No. 5. Pp. 26–34.
17. *Udovichenko S.Yu.*, *Pisarev A.D.*, *Busygin A.N.*, *Maevsky O.V.* Neuroprocessor based on combined memristor-diode crossbar // Nanoindustry. 2018. No. 5. Pp. 344–355.
18. *Kim H.K.*, *Li C.C.*, *Fang X.M.*, *Solomon J.* et al. Erbium doped semiconductor thin films prepared by RF magnetron sputtering // Materials Research Society Symposia Proceedings. 1993. Vol. 301. Pp. 55–60.
19. *Negrov D.V.*, *Kirtaeva R.V.*, *Kiseleva I.V.* et al. Integration of functional elements of resistive nonvolative memory with 1T–1R topology // Russian Microelectronics. 2016. Vol. 46. No. 6. Pp. 383–395.
20. *Pisarev A.D.*, *Busygin A.N.*, *Bobylev A.N.*, *Udovichenko S.Yu.* High element integration in logical and memory matrices of neuroprocessor by applying composite memristor-diode crossbar // International journal of nanotechnology. 2019. Vol. 16 No. 1/2/3. Pp. 182–186.
21. *Vinet M.*, *Batude P.*, *Tabone C.* et al. 3D monolithic integration: Technological challenges and electrical results // Microelectronic Engineering. 2011. V. 88. No. 4. Pp. 331–335.
22. *Zantye D.B.*, *Kumar A.*, *Sikder A.K.* Chemical mechanical planarization for microelectronics applications, // Materials Science and Engineering. 2004. Vol. 45. No. 3–6. Pp. 89–220.
23. *Or-Bach Z.*, *Wurman Z.* Integrated circuit with logic 3D. 2013. US Patent No. 8492886 B2.
24. *Deng Y.*, *Peng H.*, *Chen B.* RRAM Crossbar array with cell selection device: A device and circuit interaction study // IEEE Transactions on Electron Devices. 2013. Vol. 60. No. 2. Pp. 719–726.
25. *Писарев А.Д.*, *Маевский О.В.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* Многослойная логическая матрица на основе мемристорной коммутационной ячейки. 2019. Патент № 2682548.

26. *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. Vol. 74. No. 6. Pp. 2484–2496.
27. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor — diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
28. *Писарев А.Д.* Spice-моделирование процессов ассоциативного самообучения и безусловного разобучения в логическом блоке нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2018. № 3. С. 132–145.

# ГЛАВА 5

# ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ ВО ВХОДНОМ И ВЫХОДНОМ УСТРОЙСТВАХ БИОМОРФНОГО НЕЙРОПРОЦЕССОРА

Входное устройство нейропроцессора предназначено для первичной обработки информации в виде сжатия, фильтрации и кодирования в импульсы, в том числе биоморфные, для последующей передачи на запоминающую матрицу. Выходной блок осуществляет преобразование информации об активации нейронов в цифровой двоичный код, ее сжатие и передачу на интерфейсный блок.

Предполагается, что видеосигнал, передаваемый с интерфейсного блока на входное устройство нейропроцессора, представлен набором значений яркости пикселей в виде двоичных чисел, передаваемых по шине USB от компьютера или цифровой видеокамеры. Аудиосигнал передается аналогично, но представлен значениями амплитуд дискретного косинусного преобразования в виде двоичных чисел. Косинусное преобразование может осуществляться либо во входном устройстве, либо заранее до передачи в интерфейсный блок.

Под сжатием подразумевается дискретное косинусное преобразование и отсев малых амплитуд. Импульсы могут быть любой амплитуды и длительности, в том числе подобные биоморфным импульсам мозга с длительность 1–2 мс (см. рис. 1.8).

Часть входного устройства, реализующая импульсное кодирование, может быть построена на основе логической матрицы [1; 2]. Преобразованная информация в виде импульсов поступает на часть запоминающей матрицы, которая является единым массивом синапсов нейронов, объединенных в нейронный блок нейропроцессора. Запоминающая матрица вместе с маршрутизатором на основе логической матрицы обслуживают синаптические соединения между нейронами. Каждый выходной канал входного блока нейропроцессора представляет собой выход виртуального входного нейрона, преобразующего сенсорную информацию в импульсы. Однослойные нейросети практически не используются, гораздо более распространены нейросети с двумя и более слоями нейронов, позволяющие реализовать

более сложные зависимости выходного сигнала от входных данных. Во втором случае входные нейроны составляют половину от общего количества, поэтому количество выходных каналов кодирующей части входного блока будет равно половине от количества входов запоминающей матрицы. Учитывая, что максимальный размер запоминающей матрицы составляет $1000^2$ ячеек [3], во входном блоке необходимо использовать сверхбольшую логическую матрицу.

Далее рассмотрим, как во входном блоке значения яркости пикселей в строке видеокадра с помощью дискретного косинусного преобразования разлагаются в амплитуды гармоник. Гармоники с малой амплитудой фильтруются, а остальные с помощью перевода цифрового двоичного кода в пространственный код преобразуются в формат биоморфных импульсов.

## 5.1. ДИСКРЕТНОЕ КОСИНУСНОЕ ПРЕОБРАЗОВАНИЕ ДЛЯ ПЕРВИЧНОЙ ОБРАБОТКИ СИГНАЛОВ

Почти все современные алгоритмы сжатия с потерями (JPEG, MPEG) основаны на дискретном косинусном преобразовании DCT (Discrete Cosine Transform), являющимся разновидностью методов Фурье анализа [4]. Кроме сжатия входных данных метод DCT может быть использован в задачах: нахождения периодических закономерностей, кодирования, распознавания информации, удаления шумов и помех из информационных сигналов.

### 5.1.1. Метод дискретного косинусного преобразования

Для преобразования необходимы данные в виде отсчетов, которые обычно формируют из входных аналоговых сигналов путем дискретизации по времени и квантования, или создают искусственно. В этих данных может содержаться любая информация, в том числе яркость пикселей некоторого видеоизображения. Отсчеты группируют в виде $N$-мерного (обычно $N = 8$) входного вектора. Само DCT выполняется путем умножением входного вектора $\boldsymbol{X}$ на тензор преобразования $\hat{M}$ по формуле:

$$\boldsymbol{Y} = \hat{M}\boldsymbol{X}. \tag{5.1}$$

Результатом является $N$-мерный вектор спектра $\boldsymbol{Y}$, содержащий компоненты, соответствующие значениям амплитуд косинусных гармоник.

Наличие амплитуд с малыми значениями свидетельствуют о том, что в исходных данных отсутствует соответствующая периодичность. Процесс сжатия выполняется обнулением малых компонент спектрального вектора.

Возможен также обратный расчет выходного вектора данных по спектральным компонентам. Он выполняется с обратным тензором преобразования $\hat{M}^{-1}$ аналогично формуле (5.1). Поскольку тензор преобразования $\hat{M}$

является квадратным $N \times N$-мерным ортогональным базисом, то его определитель равен единице $|\hat{M}| = 1$ и обратный тензор преобразования равен транспонированному $\hat{M}^{-1} = \hat{M}^T$.

Существует несколько типов DCT, самый распространенный из них восьмимерный DCT-$2_8$. Компоненты тензора преобразования $M_{i,j}$ вычисляются по формуле [4]:

$$M_{i,j} = \lambda_i \cos\left(\frac{\pi}{2N}(2j+1)i\right); \quad \lambda_i = \begin{cases} \sqrt{\frac{1}{N}}, \text{ при } i = 0; \\ \sqrt{\frac{2}{N}}, \text{ при } i \neq 0, \end{cases} \tag{5.2}$$

где $N$ — размер тензора (равен 8 для DCT-$2_8$); $i$ и $j$ — индексы строк и столбцов, изменяющиеся от 0 до $N - 1$; $\lambda_i$ — нормирующий множитель.

Для демонстрации принципа сжатия с последующей фильтрацией и поиска закономерностей во входных данных на рис. 5.1 показаны результаты DCT 8-битного прямого и обратного преобразований монохромных фрагментов тестового изображения [5]. Изображения представлены на графике полосами сверху и снизу. Изображение можно преобразовывать достаточно большими участками, но обычно для сжатия его делят на фрагменты по 8 пикселей, как и в настоящем примере. Входной вектор $\boldsymbol{X}$ определяется значениями яркостей пикселей фрагмента. График компонент $\boldsymbol{X}$ показан на рис. 5.1, *а*. Входной вектор состоит из восьми чисел 8-битовой разрядности (целые числа, изменяющиеся в диапазоне от 0 до 255). Понижение уровня в центре графика соответствует темной области в центре фрагмента изображения.

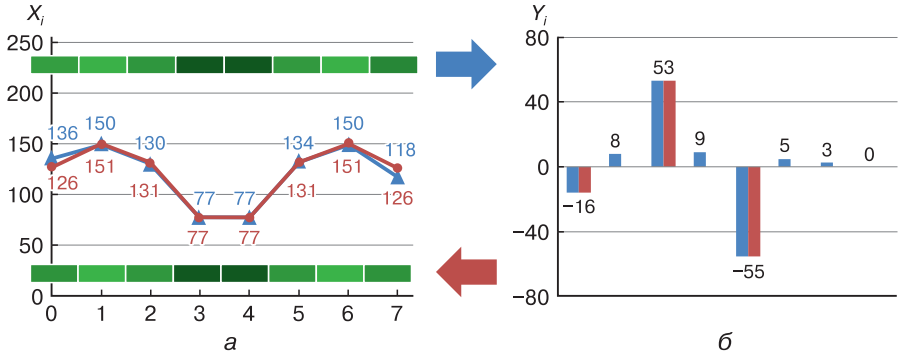



**Рис. 5.1.** Прямое (синий цвет) и обратное (красный цвет) DCT-$2_8$:
*а* — фрагменты тестового изображения, график зависимости яркости от номера пикселя; *б* — диаграмма спектральных амплитуд

Спектральный вектор $\boldsymbol{Y}$, рассчитанный по формуле (5.1), показан на рис. 5.1, *б*. Для расчета входному вектору было дано смещение −127, чтобы значения компонент находились в диапазоне от −127 до 128. На каждом

этапе результат расчета округлялся до целого числа. На диаграмме заметны три компонента с наибольшим значением, которые заключают в себе основную часть энергии исходного сигнала.

Для сжатия 5 компонент с малыми амплитудами обнуляли, затем выполняли обратное преобразование. Результат преобразования сжатого спектрального вектора показан на рис. 5.1, *а*. Восстановленный вектор данных по трем спектральным компонентам из восьми отражает входной сигнал, но с небольшими отклонениями рассчитанной яркости пикселей от исходных значений. Основная информация на полученном фрагменте изображения в виде темной полосы не потеряна. Полученный фрагмент изображения стал симметричным относительно своей центральной части. Потеря информации при DCT приводит к симметрии из-за разложения по гармоническому базису.

### 5.1.2. Быстрый алгоритм дискретного косинусного преобразования для входного блока нейропроцессора

Расчет DCT по классической формуле требует большого количества операций двух типов — умножения и сложения чисел в формате с плавающей точкой. Операция умножения требует во много раз больше элементарных преобразований на аппаратном средстве, чем операция сложения. Количество элементарных преобразований, выполняемых в аппаратном средстве, определяет сложность метода вычислений.

Из-за высокой сложности неадаптированного алгоритма DCT существует проблема низкой скорости работы устройств. Действительно, при компрессии видеопотоков требуется делать преобразования максимально быстро. Проблема низкой скорости преобразования решается двумя путями. Во-первых, разрабатываются специализированные процессоры и сопроцессоры, ориентированные на скоростное вычисление тензорных операций. Во-вторых, совершенствуются алгоритмы, которые уменьшают сложность расчета DCT за счет сокращения повторяющихся операций умножения и сложения чисел с плавающей точкой.

Из множества существующих быстрых алгоритмов DCT, описанных в [4], для применений во входном блоке нейропроцессора наиболее подходящим может быть алгоритм, учитывающий связи между алгебраическими свойствами значений базисных функций DCT и структурой тензора преобразования $\hat{M}$.

В табл. 5.1 представлены рассчитанные числовые значения тензора преобразования по формуле (5.2). Значения во многих ячейках табл. 5.1 совпадают, некоторые с точностью до знака. Ячейки, имеющие одинаковые значения, выделены цветом. Значения, отличающиеся только знаком, отмечены разной яркостью. Всего получилось 7 уникальных цветов без учета знака. Таким образом, спектральные компоненты выходного вектора $\boldsymbol{Y}$

выражаются через суперпозицию 7 вариантов произведений на постоянные числа компонент входного вектора.

*Таблица 5.1*

**Числовые компоненты тензора $\hat{M}$ преобразования DCT-$2_8$**

| i \ j | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| **0** | 0,354 | 0,354 | 0,354 | 0,354 | 0,354 | 0,354 | 0,354 | 0,354 |
| **1** | 0,49 | 0,416 | 0,278 | 0,098 | –0,098 | –0,278 | –0,416 | –0,49 |
| **2** | 0,462 | 0,191 | –0,191 | –0,462 | –0,462 | –0,191 | 0,191 | 0,462 |
| **3** | 0,416 | –0,098 | –0,49 | –0,278 | 0,278 | 0,49 | 0,098 | –0,416 |
| **4** | 0,354 | –0,354 | –0,354 | 0,354 | 0,354 | –0,354 | –0,354 | 0,354 |
| **5** | 0,278 | –0,49 | 0,098 | 0,416 | –0,416 | –0,098 | 0,49 | –0,278 |
| **6** | 0,191 | –0,462 | 0,462 | –0,191 | –0,191 | 0,462 | –0,462 | 0,191 |
| **7** | 0,098 | –0,278 | 0,416 | –0,49 | 0,49 | –0,416 | 0,278 | –0,098 |

Все 7 чисел в разном порядке находятся в каждом столбце тензора преобразования (см. табл. 5.1). В соответствии с выражением (5.2) все уникальные значения в ячейках вычисляются по формуле:

$$M_{i,\,0} = \frac{1}{2}\cos\left(\frac{\pi}{16}i\right), \tag{5.3}$$

где $i$ — индекс числа, изменяющийся от 1 до 7.

Используя простые тригонометрические зависимости для косинуса кратных углов $\cos(2\varphi) = 2\cos^2(\varphi) - 1$, $\cos(3\varphi) = -3\cos(\varphi) + 4\cos^3(\varphi)$, можно выразить косинусы в (3) через $\cos(\pi/16)$. Полученные зависимости будут представлять собой многочлены Чебышева первого рода. Если ввести обозначение $\cos(\pi/16) = \zeta$, то для компонент тензора DCT — чисел $M_{i,\,0}$ можно записать формулы

$$\begin{gathered}
M_{1,\,0} = \frac{1}{2}\zeta; \quad M_{2,\,0} = \frac{1}{2}\left(2\zeta^2 - 1\right); \\
M_{3,\,0} = \frac{1}{2}\left(4\zeta^3 - 3\zeta\right); \quad M_{4,\,0} = \frac{1}{2}\left(8\zeta^4 - 8\zeta^2 + 1\right); \\
M_{5,\,0} = \frac{1}{2}\left(16\zeta^5 - 20\zeta^3 + 5\zeta\right); \quad M_{6,\,0} = \frac{1}{2}\left(32\zeta^6 - 48\zeta^4 + 18\zeta^2 - 1\right); \\
M_{7,\,0} = \frac{1}{2}\left(64\zeta^7 - 112\zeta^5 + 56\zeta^3 - 7\zeta\right).
\end{gathered} \tag{5.4}$$

Многочлен Чебышева первого рода можно рассматривать как представление числа в системе счисления с иррациональным косинусным основанием

$$\zeta = \cos\left(\frac{\pi}{16}\right) = \frac{1}{2}\sqrt{2+\sqrt{2+\sqrt{2}}} \approx 0{,}980785, \tag{5.5}$$

в которой коэффициенты многочлена седьмой степени вида

$$f = a_0 + a_1\zeta + a_2\zeta^2 + a_3\zeta^3 + a_4\zeta^4 + a_5\zeta^5 + a_6\zeta^6 + a_7\zeta^7$$

являются разрядами (цифрами) этой системы счисления. В этом случае числа тензора преобразования представить следующим образом

$$\begin{aligned}
2M_{1,\,0} &= .0.0.0.0.0.0.1.0_{\zeta};\\
2M_{2,\,0} &= .0.0.0.0.0.2.0.-1_{\zeta};\\
2M_{3,\,0} &= .0.0.0.0.4.0.-3.0_{\zeta};\\
2M_{4,\,0} &= .0.0.0.8.0.-8.0.1_{\zeta};\\
2M_{5,\,0} &= .0.0.16.0.-20.0.5.0_{x};\\
2M_{6,\,0} &= .0.32.0.-48.0.18.0.-1_{\zeta};\\
2M_{7,\,0} &= .64.0.-112.0.56.0.-7.0_{\zeta}.
\end{aligned} \tag{5.6}$$

Точками в (5.6) отделены разряды числа, младший разряд записан справа. Если заменить основание системы счисления по формуле

$$\nu = 2\zeta = 2\cos\left(\frac{\pi}{16}\right) = \sqrt{2+\sqrt{2+\sqrt{2}}} \approx 1{,}961570, \tag{5.7}$$

то можно уменьшить числовые значения разрядов и оставить их целыми:

$$\begin{aligned}
4M_{1,\,0} &= .0.0.0.0.0.0.1.0_{\nu};\\
4M_{2,\,0} &= .0.0.0.0.0.1.0.-2_{\nu};\\
4M_{3,\,0} &= .0.0.0.0.1.0.-3.0_{\nu};\\
4M_{4,\,0} &= .0.0.0.1.0.-4.0.2_{\nu};\\
4M_{5,\,0} &= .0.0.1.0.-5.0.5.0_{\nu};\\
4M_{6,\,0} &= .0.1.0.-6.0.9.0.-2_{\nu};\\
4M_{7,0} &= .1.0.-7.0.14.0.-7.0_{\nu}.
\end{aligned} \tag{5.8}$$

Аппаратная реализация упрощается, если расчет следующего произведения производить по разрядам предыдущего, воспользовавшись рекуррентным соотношением, которое следует из (5.8):

$$M_{1,\,0}(\nu) = \nu;$$

$$M_{2,\,0}(\nu) = \nu^2 - 2; \tag{5.9}$$

$$M_{i+1,\,0}(\nu) = \nu M_{i,\,0}(\nu) - M_{i-1,\,0}(\nu).$$

Принимая во внимание формулы (5.8) и (5.9), в аппаратном средстве, реализующем расчеты в предложенной системе счисления, требуется выполнять операции: перестановку разрядов числа для умножения на *н* и вычитание целых чисел, являющихся разрядами сложного числа в предложенном формате счисления.

Таким образом, алгоритм вычисления DCT упрощается. Для этого требуется определить $7 \times 8 = 56$ уникальных произведений в предложенной системе счисления, выражающей числа с помощью многочленов Чебышева, а затем их суммировать или вычитать в порядке, отмеченном цветами чисел в строках табл. 1. Преимущество использования многочленов Чебышева в расчете DCT аппаратным средством заключается в том, что операции умножения и сложения проводятся с целыми числами. С учетом рекуррентного соотношения метод DCT в аппаратном средстве можно привести к операциям перестановки, суммирования и вычитания целых чисел.

В конце всех основных вычислительных операций результат DCT можно анализировать в полученной системе счисления, и без каких-либо преобразований передавать его в основное ядро нейропроцессора для следующего этапа обработки.

При необходимости можно выполнить перевод результата DCT в более информативную систему счисления, для этого понадобятся вычисления чисел в формате «плавающей точки». Выполнение DCT с целью сжатия информации с потерями потребует осуществить операции над числами с «плавающей точкой» не со всеми спектральными компонентами, при этом допускается ограничение точности результата.

### 5.1.3. Адаптация быстрого алгоритма дискретного косинусного преобразования к входному блоку нейропроцессора

Адаптации алгоритма быстрого DCT к входному блоку нейропроцессора заключается в представлении карты коммутируемых мемристорных связей, передающих информационные импульсы, между пластами сверхбольшой 3D логической матрицы. Вследствие громоздкости и большой сложности пластов и карт связей сверхбольшой 3D-матрицы информационные потоки нагляднее всего показывать с помощью ориентированного графа.

Вершинами графа являются промежуточные целочисленные значения параметров преобразования. В ребрах графа представлены направления передачи и простые операции параметров преобразования. Представленные графы показывают для аппаратного средства принципы разложения сложного DCT на простые операции с целочисленными значениями.

Для решения данной задачи можно предложить несколько вариантов графа, среди которых условно можно выделить две крайности. Первая содержит подход с экономией памяти и состоит из наименьшего количества вершин, сохраняющих переменные, но предполагает много сложновычисляемых связей. Ко второй части следует отнести вариант, содержащий большое количество вершин, которые объединены простыми зависимостями. Как правило, первый подход, характеризуемый низкой скоростью работы, представляет меньший практический интерес для аппаратной реализации по сравнению со вторым, отличающимся быстродействием, но требующим большего количества вершин и ребер.

На рис. 5.2 предложен один из быстрых вариантов графов реализации DCT по векторно-тензорной формуле (5.1) в электронном устройстве на основе 3D-матрицы. Матрица во входном блоке нейропроцессора запрограммирована только на целочисленные операции сложения, вычитания и перестановки позиций чисел, представленных в системе счисления с основанием

$$\nu = 2\cos\left(\frac{\pi}{16}\right).$$

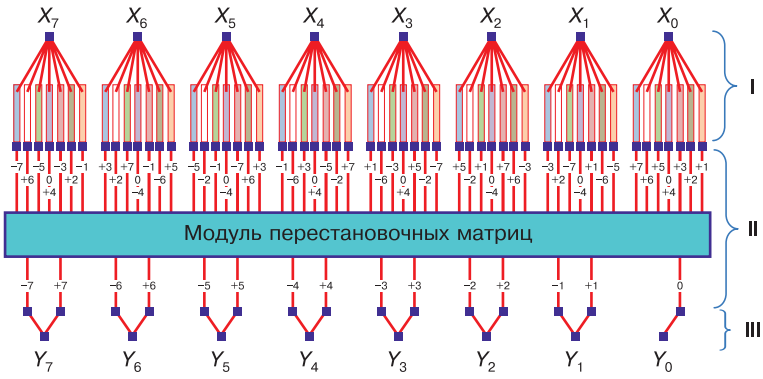



**Рис. 5.2.** Граф реализации быстрого DCT во входном блоке нейропроцессора

Граф состоит из трех слоев, между которыми осуществляется передача информации сверху вниз. В первом слое (область I на рис. 5.2) показаны 8 узлов входного вектора $\boldsymbol{X}$, разделенных на $7 \times 8 = 56$ ребер, отвечающих за произведения компонент вектора на числа матрицы преобразования $\hat{M}$.

Во втором слое (область II на рис. 5.2) с помощью модуля перестановочных матриц выполняется перенаправление результатов произведения и их соответствующее суммирование. При этом положительные и отрицательные члены компонент выходного вектора $\boldsymbol{Y}$ разделяются на два потока. В третьем слое (область III на рис. 5.2) представлены 8 вершин значений выходного вектора $\boldsymbol{Y}$, которые получаются путем вычитания результатов перестановочной матрицы предыдущего слоя.

На рис. 5.3 представлен рекуррентный способ целочисленного вычисления произведений чисел тензора $\hat{\boldsymbol{M}}$ на компоненты входного вектора $\boldsymbol{X}$.

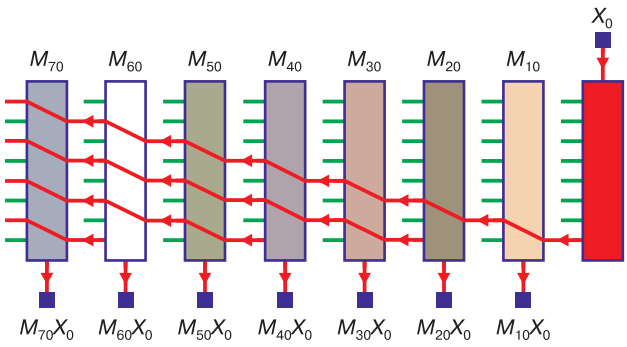



**Рис. 5.3.** Рекуррентная реализация произведений DCT в системе счисления с иррациональным косинусным основанием

Эта операция показана в первом слое (область I на рис. 5.2) графа реализации быстрого DCT. Способ приводит сложные вычисления к целочисленным операциям, одинаково выполняемым для всех компонент входного вектора. Способ основывается на представлении иррациональных чисел в целочисленном виде с помощью системы счисления с иррациональным косинусным основанием, выраженным формулой (5.7).

Представленные в (5.8) разряды компонент тензора преобразования показаны на рис. 5.3 в виде восьми зеленых линий. Ненулевые значения разрядов отмечены красным цветом. Нахождение произведений осуществляется в соответствии с рекуррентным соотношением (5.9), в котором каждое вычисляемое произведение определяются по двум предыдущим. Выполнение операции сдвига входной величины и вычитания разрядов предыдущего числа показаны на рис. 5.3 красными линиями.

Суммирование и вычитание произведений осуществляется во втором слое (область II на рис. 5.2) представленного графа. В этом модуле основной операцией являются перестановки значений произведений для подачи на сумматоры. Сумматоров два вида: для положительных и отрицательных произведений, как показано в третьем слое графа (область III на рис. 5.2).

Для определения порядка произведений в модуле перестановочных матриц представлена табл. 5.2, полученная из табл. 5.1.

*Таблица 5.2*

**Порядок индексов компонент векторов и тензора для использования в перестановочных операциях быстрого DCT алгоритма**

| *i* \ *j* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| **0** | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| **1** | 1 | 3 | 5 | 7 | –7 | –5 | –3 | –1 |
| **2** | 2 | 6 | –6 | –2 | –2 | –6 | 6 | 2 |
| **3** | 3 | –7 | –1 | –5 | 5 | 1 | 7 | –3 |
| **4** | 4 | –4 | –4 | 4 | 4 | –4 | –4 | 4 |
| **5** | 5 | –1 | 7 | 3 | –3 | –7 | 1 | –5 |
| **6** | 6 | –2 | 2 | –6 | –6 | 2 | –2 | 6 |
| **7** | 7 | –5 | 3 | –1 | 1 | –3 | 5 | –7 |

Одинаковые произведения отмечены соответствующими цветами и цифрами в ячейке. В табл. 5.2 связаны три величины: номер столбца, номер строки и цифра ячейки. Номер столбца таблицы задает индекс компоненты выходного вектора ***Y***. Номер строки определяет индекс группы сумм, соответствующих компонентам входного вектора ***X***. Значение ячейки определяет номер произведения в группе. Знак ячейки задает отрицательное или положительное направление суммирования в вершинах графа, выполняемое на границе второго и первого слоев (области II и III на рис. 5.2). Перестановка произведений отмечена соответствующими числами, показанными на линиях связи второго слоя (область II на рис. 5.2).

Алгоритм DCT, адаптированный к 3D логической матрице, имеет высокие быстродействие и энергоэффективность. В быстром алгоритме DCT выполняется последовательность простых операций: суммирования, вычитания и перестановки целых чисел. Применяется точное представление иррациональных чисел с помощью целочисленных разрядов системы счисления с иррациональным косинусным основанием. Результаты DCT представляются в выбранной системе счисления целыми разрядами.

При адаптации DCT к входному блоку нейропроцессора используется тот факт, что тензор преобразования содержит только 7 независимых компонент из 64. Реализация умножения вектора на тензор производится аналогичными модулями 3D логической матрицы. Представленное рекуррентное соотношение между разными семью компонентами тензора преобразования позволяет значительно упростить определения произведений DCT. Возможность реализации перестановочных операций в 3D логической матрице используется для вычисления положительных и отрицательных частей компонент выходного вектора.

Увеличение скорости DCT в сверхбольшой 3D логической матрице достигается за счет применения простых операций, выполняемых параллельно в логических связанных блоках. Скорость работы такой системы может быть крайне высока и определяется временем одного тактового импульса, ограниченного лишь скоростью срабатывания инверторных элементов и распространением сигналов по шинам комбинированного мемристоро-диодного кроссбара 3D логической матрицы.

Высокая энергоэффективность дискретного косинусного преобразования в 3D логической матрице достигается за счет распределения в пространстве формирующих сигнал элементов схемы при их ключевой работе. По сравнению с известным аналоговым матрично-векторным умножителем Hewlett-Packard (1T1M) [6] логическая матрица (1D1M) [1; 2] при малой размерности входного вектора и, соответственно, матрицы преобразования является более энергоэффективной. Однако с ростом размерности входного вектора число элементов в матрице 1D1M возрастает экспоненциально, а в аналоговой — квадратично. На рис. 5.4 показана зависимость количества ячеек в указанных аналоговом и цифровых массивах при точности чисел 6 бит от числа компонент во входном векторе.

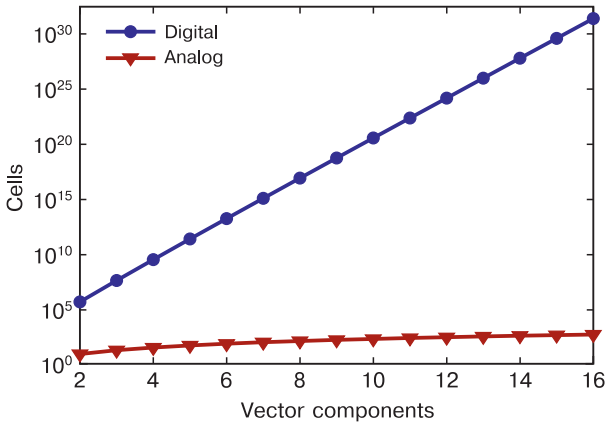



**Рис. 5.4.** Число ячеек цифровой и аналоговой матриц в зависимости от числа 6-битных компонент входного вектора (пикселей)

Несмотря на то что аналоговые вычисления в массиве НР являются энергозатратными и дают искажения выходного вектора до 2 %, на основании проведенного анализа можно сделать выбор в пользу этого массива при сжатии информации с помощью дискретного косинусного преобразования во входном блоке нейропроцессора. Для обслуживания большой запоминающей матрицы нейропроцессора (500 входов) можно задействовать параллельно несколько таких массивов. При точности вычислений 6 бит максимальный размер аналогового массива НР составляет 64 × 64 ячеек.

## 5.2. БИОМОРФНОЕ ИМПУЛЬСНОЕ КОДИРОВАНИЕ ИНФОРМАЦИИ В ЭЛЕКТРОННЫХ НЕЙРОНАХ, РЕАЛИЗУЕМЫХ НА БАЗЕ ЭЛЕМЕНТОВ ЛОГИЧЕСКОЙ МАТРИЦЫ

### 5.2.1. Принципы импульсного кодирования информации в биологических системах

В основе импульсного кодирования во входном блоке нейропроцессора с целью достижения высокой энергоэффективности предлагается использовать известные принципы информационной работы биологических сенсорных нейронов. Стоит отметить, что многие аспекты работы биологической сенсорной системы еще не получили точного научного объяснения. Однако достоверно установлены некоторые принципы, являющиеся достаточно простыми и информационно-функциональными. Эти принципы рассмотрим с точки зрения применения их во входном блоке нейропроцессора.

Известно, что в биологической системе информация об особенностях поступающих входных стимулов содержится в последовательности возбуждений электрохимической природы, называемых в нейробиологии потенциалом действия, исследованного впервые в работах Hodgkina-Huxley [8]. Потенциал действия описывается достаточно сложной системой дифференциальных уравнений для изменения электрического потенциала во времени и в пространстве вдоль нейронального волокна. Форма зависимости электрического потенциала является следствием физиологической работы живой клетки нейрона, перераспределяющей волнообразно градиент концентрации ионов на своей мембране за счет ионных насосов и электрического поля. Пиковая часть потенциала действия называется спайком (*spike*) [9]. Считается, что спайк является основным информационным носителем в биологических нейронных сетях.

Для воплощения в нейропроцессоре биоморфного механизма передачи информации эти факты используются следующим образом. Аналог спайка может быть эффективным носителем информации во входном блоке нейропроцессора, если он будет реализован в виде короткого электрического импульса без учета сложных особенностей формы всего потенциала действия. В этом случае ионные наносы эквивалентны источнику электрического питания, подведенного при помощи электрических шин к узлам электрической сети микросхемы.

Существует научное подтверждение того, что форма потенциала действия участвует в формировании следа памяти, например, по механизму STDP (*spike-timing-dependent plasticity* — пластичность, зависящая от времени спайка), описанного в работе [9]. Однако во входном устройстве нейропроцессора не требуется реализация механизмов памяти, поэтому процесс STDP может не учитываться при разработке способа кодирования входной

информации. Поскольку нет других данных, что временные, амплитудные или другие характеристики самих импульсов кодируют информацию, можно ограничить модель входного блока нейропроцессора тем фактом, что информация будет заключена только в моментах времени появления электрических импульсов, имитирующих спайки, генерируемых сенсорными нейронами.

Кодирование информации стимула обычно осуществляется по принципу — чем больше уровень стимула, тем выше частота спайков. Проведенные в последнее время исследования показывают, что информация кодируется сенсорными нейронами не только частотой генерирования спайков, но и относительным моментом времени их появления. Этот механизм передачи информации назван spike-timing [10]. Spike-timing как биологический механизм информационного кодирования имеет качественное преимущество перед механизмом, основанным на прямой зависимости частоты спайков от аналогового уровня входного стимула. Преимущество заключается в возможности наиболее быстро передавать в центральную нервную систему изменения в стимулах, оказываемые на сенсорные нейроны. Это подтверждается быстротой передачи тактильных событий с кончика пальца человека во время манипуляций с объектом [11], или исследованиями по прохождению информационных сигналов от нейронов сетчатки глаза при кратко проецируемых на нее изображениях [12]. Выяснено, что в этих случаях в биологической нейронной системе спайки генерируются с высокой временной точностью, в результате чего задержка в прохождении информационного сигнала обычно составляет не больше нескольких миллисекунд при переменных условиях [13–15].

Есть значительное отличие импульса, генерируемого в электронных системах, от спайка биологической системы. В отличие от электрических импульсов спайки распространяются по нейрональным волокнам с относительно невысокой скоростью (в среднем от 1 до 100 м/с в зависимости от размера волокна и наличия миелинизированной оболочки [16]). Спайки распространяются по нейрональному волокну, перенося, таким образом, сигналы входной информации к нейронам центральной нервной системы с разными задержками между сенсорными нейронами и нейронами центральной нервной системы. Очевидно, что задержка во всей системе — это величина переменная, которая зависит от размера, длины нейронного волокна и наличия миелинизированной оболочки. Эти задержки могут играть основную роль в энергоэффективном кодировании информации.

Новые нейрофизиологические результаты [17] показывают, что обработка информации в нейронных системах может основываться на точной синхронизации потенциалов действия во всей нейронной системе. В этом случае информация содержится в совокупности поступающих спайков, собираемых в пространстве с фиксированными задержками по времени от сенсорных нейронов. Образующуюся совокупность спайков называют временным паттерном [10; 18]. В каждом временном паттерне спайки,

приходящие от разных сенсорных нейронов, занимают некоторое точное положение. Есть основания полагать, что головной мозг воспринимает информационный поток, различая повторяющиеся временные паттерны спайков. Этот биологический информационный механизм может оказаться очень энергоэффективным в случае его реализации во входном блоке нейропроцессора, из-за минимизации количества импульсов, приходящихся на передачу единичного объема информации.

Важность положения спайка как средства передачи информации простимулировала ряд исследований спайковых нейронных сетей SNN (Spiking Neural Network) [19; 20]. SNN являются биоморфными нейронными сетями третьего поколения, которые наиболее реалистично на сегодняшний день отражают работу биологической нейронной системы. Предложенные правила SPAN [21] и PSD [22] продемонстрировали успех в обучении SNN при формировании представлений пространственно-временных моделей спайковых паттернов. SNN может являться моделью построения импульсной работы всего нейропроцессора.

Для воплощения биологических временных паттернов во входном блоке нейропроцессора, ко всему прочему, требуется реализовать способ организации регулируемых временных задержек электрических импульсов. Эти импульсы передают информацию по электрическим проводникам между узлами, играющими роль нейронов. Информация во входном блоке, поступая в центральные блоки нейропорцессора, проходит через несколько слоев входных нейронов, где осуществляется кодирование сигналов. В качестве информационной ячейки электронного нейрона в данном случае можно рассматривать электрическую шину, на которой осуществляется суммация поступающих импульсов (термин «суммация» заимствован из нейрофизиологии, где под ним понимается слияние эффектов ряда стимулов).

Резюмируя представленные аспекты работы биологических нейронных систем, можно выделить следующие факты, используемые для разработки входного блока нейропроцессора. В биологической нервной системе информация кодируется временными паттернами последовательности спайков, генерируемых нейронами при их возбуждении. Спайк представляет собой пик электрического потенциала — возмущение, которое распространяется по нервным волокнам с небольшими скоростями, перенося единичными актами информацию между нейронами. Временной паттерн представляет собой группу спайков с точным пространственно-временным расположением, которое возникает как реакция на определенное входное событие. Увеличение уровня сигнала во входной информации кодируется повышением частоты генерируемых спайков сенсорными нейронами, передаваемых далее для обработки в центральные отделы нервной системы. Информация в биологической нейронной системе, перед поступлением в центральные отделы, фильтруется от помех и шумов, подвергается сжатию и может многократно перекодироваться.

Таким образом, спайки могут быть реализованы во входном устройстве нейропроцессора в виде электрических импульсов. Информационные параметры входного сигнала рациональнее всего в мемристорных матрицах кодировать частотой и пространственным распределением сигналов. Для имитации биоморфного механизма временных паттернов спайков требуется реализация регулируемых задержек при распространении электрических импульсов по проводникам между узлами, воспроизводящими функциональность нейронов. В группе таких узлов одновременно с частотой импульсов и их пространственным распределением можно кодировать информацию за счет точных моментов времени появления электрических импульсов, формируя тем самым картины импульсных последовательностей похожих на временные паттерны спайков в биологических нейронных системах. Такие импульсные последовательности будут поступать с входного блока в следующие блоки нейропроцессора для их дальнейшей обработки. Далее будет показано, как представленные биоморфные принципы могут способствовать высокой энергоэффективности нейропроцессора в случае воплощения их во входном блоке.

### 5.2.2. Схема и принцип работы электронного нейрона, реализуемого на базе элементов 3D логической матрицы

Высокая энергоэффективность может быть достигнута за счет применения в нейропроцессоре электронного аналога нейрона, биоморфно кодирующего входную информацию. Передача информации между электронными нейронами может выполняться путем имитирования механизма биологических временных спайковых паттернов.

На рис. 5.5, *а* представлена оригинальная схема электронного нейрона [23]. Электронный нейрон построен с использованием минимального количества элементов. Так сделано для того, чтобы простые электронные схемы, имитирующие нейроны, можно было объединять в сверхбольшие массивы и реализовать в объеме 3D логической матрицы. На рис. 5.1, *а* показаны функциональные блоки классического информационного нейрона, взвешивающего, суммирующего и сравнивающего результат с порогом. Ниже, на рис. 5.5, *б* показана принципиальная схема электронного нейрона, имитирующего блоки классического нейрона. Резисторы $R_1$–$R_5$ играют роль синаптических весов, образуя делитель напряжения входных импульсов с резистором $R$. На общей шине происходит суммирование втекающих токов, что подобно работе биологического нейрона. Общая шина подключена к повторителю $DD_1$, который имитирует выход аксона с пороговой функцией активации. На рис. 5.5, *в* показана передаточная характеристика повторителя, которая имеет ступенчатую форму в области порога срабатывания в середине диапазона питания, что аналогично функциональности блока аксона в информационном нейроне.

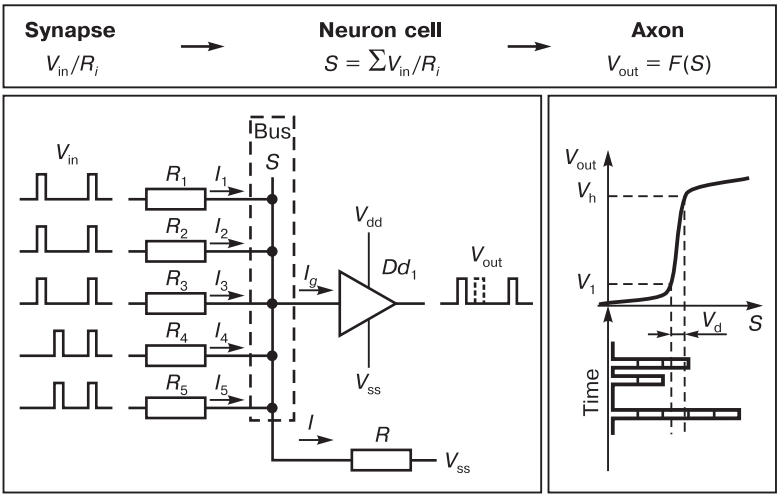



**Рис. 5.5.** Принцип суммации импульсов в электронном нейроне:

*а* — функциональная схема информационного нейрона;
*б* — принципиальная схема электронного нейрона;
*в* — передаточная характеристика повторителя $DD_1$, выполняющего пороговую функцию

В качестве примера входной информации, подаваемой на электронный нейрон, представлены некоторые два числовых значения. Они закодированы простым аналогом биоморфного временного паттерна, состоящего из двух импульсов, играющих роль спайков. Значения чисел закодированы интервалом между первым и вторым импульсом, показанным на пяти эпюрах входного напряжения $V_{in}$ (рис. 5.5, *б*). Импульсы поступают в электронный нейрон через электронные синапсы. Амплитуды импульсов напряжения суммируются на общей шине при сложении втекающих в нее импульсов тока. В случае одинаковых номиналов у резисторов амплитуда импульса на шине определяется количеством одновременно поступающих входных импульсов. При превышении порога напряжения в области d$V$ (рис. 5.5, *в*) на суммирующей шине происходит переключение выхода инвертора $DD_1$ между состояниями с напряжениями $V_l$, $V_h$ и формирование на выходе кратковременного импульса.

Таким образом, через электронный нейрон, работающий в режиме суммации импульсов, проходят только те импульсы, количество которых в данный момент времени больше заданного порога. Информация о других импульсах отфильтровывается. Прохождение групп импульсов демонстрирует принцип суммации, характерный для работы биологических нейронов. Функция определения максимального количества совпадающих временных паттернов импульсов требуется для фильтрации информационных потоков во входном блоке нейропроцессора.

### 5.2.3. Условия формирования биоморфных импульсов на шинах 3D логической матрицы

В отличие от упрощенной резистивной схемы (см. рис. 5.5), показывающей принцип работы электронного нейрона, в реальном устройстве требуется подбор параметров мемристорных слоев и селективных элементов с целью выявления условий, подходящих для суммации импульсов на шинах. В твердотельных мемристорах на основе переходных изменение сопротивления происходит в достаточно широком диапазоне от $10^3$ до $10^6$ Ом. В мемристоре наблюдается множество устойчивых состояний проводимости, которые могут играть роль аналога биологических синаптических состояний. Стоит задача физического моделирования, заключающаяся в получении максимально возможного числа связей, имитирующих синапсы нейронов.

На рис. 5.6 показана схема фрагмента 3D логической матрицы [1; 2], полная конструкция которой представлена в работе [24].

Импульсные сигналы поступают на общую шину $V_g$ через входные инверторы и мемристорные ячейки. На схеме рис. 5.6 мемристорные ячейки объединены в группы по общему режиму работы. Выделены четыре следующих режима, которые отмечены номерами, обведенными окружностями на схеме.

1 — Режим соответствует мемристорам, запрограммированным в *высокопроводящее состояние*, на которые в данный момент времени поступает *низкий уровень* напряжения. Низкий уровень напряжения подается от входного КМОП-инвертора, пропускающего через себя информационный импульс электрической шины предыдущего слоя.

2 — Режим соответствует мемристорам, запрограммированным в *низкопроводящее состояние*, на которые в данный момент времени также поступает *низкий уровень* напряжения от входного КМОП-инвертора.

3 — Режим соответствует мемристорам, запрограммированным в *высокопроводящее состояние*, на которые в данный момент времени поступает *высокий уровень* напряжения. Высокий уровень напряжения подается от входного КМОП-инвертора, который в данный момент времени не пропускает через себя информационный импульс.

4 — Режим соответствует мемристорам, запрограммированным в *низкопроводящее состояние*, на которые в данный момент времени поступает *высокий уровень* напряжения. Высокий уровень напряжения подается от входного КМОП-инвертора, который в данный момент времени не пропускает через себя информационный импульс.

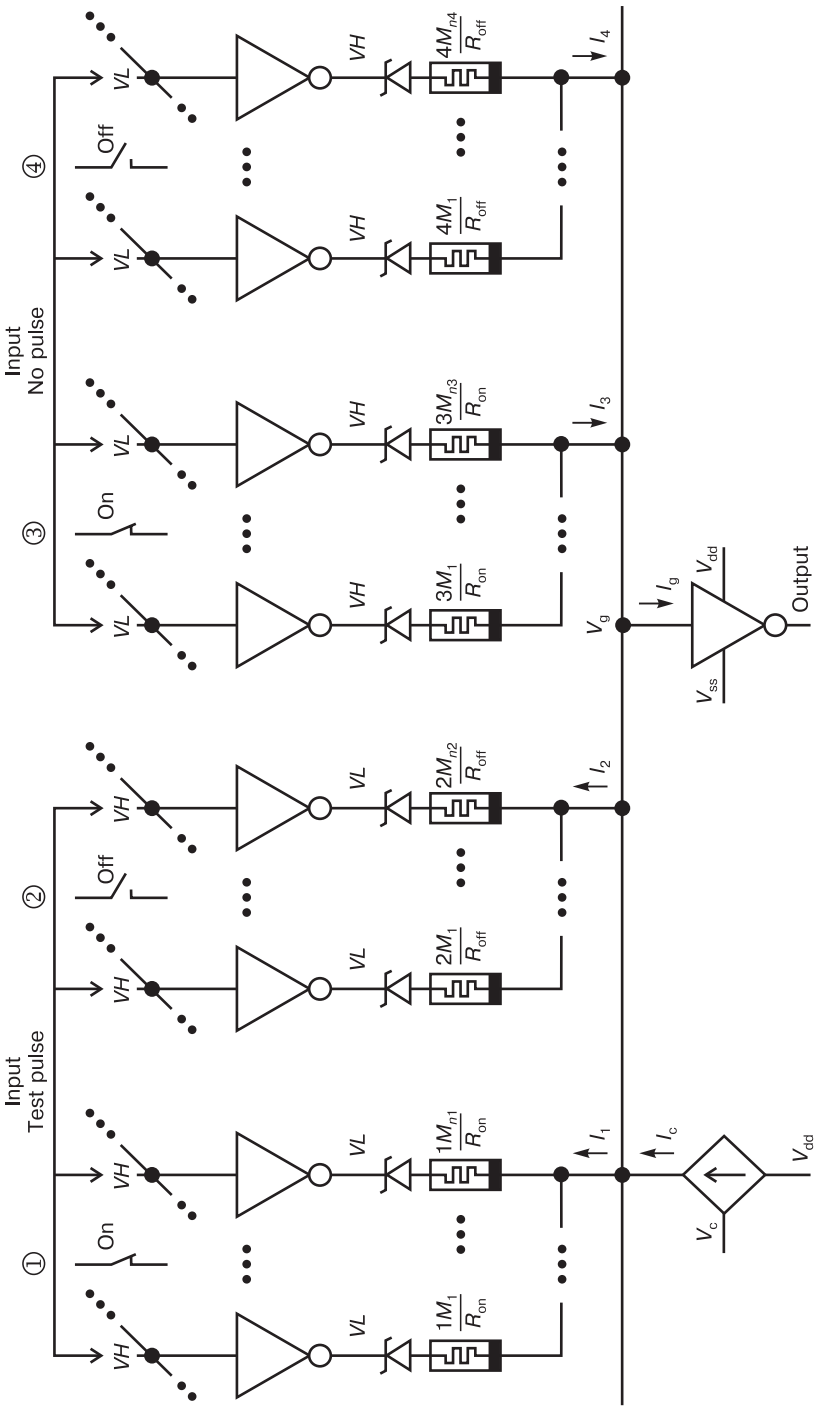



**Рис. 5.6.** Фрагмент электрической схемы 3D логической матрицы [23]

Напряжение на шине $V_g$ входит в алгебраическую сумму токов, записанную для общего узла:

$$-\sum_{i=1}^{n1}\left(\frac{V_g - V_{Li}}{R^{on}_{1M_i} + R^{+}_{z}}\right) - \sum_{i=1}^{n2}\left(\frac{V_g - V_{Li}}{R^{off}_{2M_i} + R^{+}_{z}}\right) + \sum_{i=1}^{n3}\left(\frac{V_{Hi} - V_g}{R^{on}_{3M_i} + R^{-}_{z}}\right) + \sum_{i=1}^{n4}\left(\frac{V_{Hi} - V_g}{R^{off}_{4M_i} + R^{-}_{z}}\right) + I_c = 0, \tag{5.1}$$

где $V_g$ — напряжение на шине, относительно точки нижнего потенциала $V_{ss}$; $V_{Hi}$ и $V_{Li}$ — уровни напряжения, соответствующие логическим единице и нулю, которые формируют КМОП-инверторы; $I_c$ — поступающий ток в шину от источника, установленного на периферии матрицы и предназначенного для управления порогом срабатывания КМОП-инвертора; $R^{on}_{1M_i}$, $R^{off}_{2M_i}$, $R^{on}_{3M_i}$ и $R^{off}_{4M_i}$ — резистивные состояния мемристоров разделенные по четырем группам; $R^{+}_{z}$, $R^{-}_{z}$ — сопротивления диодов Зенера в прямом и обратном направлениях прохождения тока; $n_1$, $n_2$, $n_3$ и $n_4$ — полное количество мемристоров в группах с одинаковым режимом работы.

В сумме (5.1) первые четыре компонента соответствуют группам мемристоров, работающих в одинаковых условиях, показанных на схеме (см. рис. 5.6). Эти компоненты можно упростить, сняв знаки суммирования, предполагая, что логические уровни инверторов, разброс мемристорных резистивных состояний и сопротивлений диодов Зенера не вносят значительную ошибку в формировании импульса выходного инвертора. Это допустимо еще и потому, что в нейроморфной системе не требуется полная точность прохождения всех сигналов. Поэтому в формуле (5.1) можно снять суммирующие операторы с учетом введенных обозначений, определяемых параметрами 3D логической матрицы

$$R^{on}_{1M_i} + R^{+}_{z} = \frac{1}{\sigma^{+}_{on}}; \quad R^{off}_{2M_i} + R^{+}_{z} = \frac{1}{\sigma^{+}_{off}}; \quad R^{on}_{3M_i} + R^{-}_{z} = \frac{1}{\sigma^{-}_{on}};$$

$$R^{off}_{4M_i} + R^{-}_{z} = \frac{1}{\sigma^{-}_{off}}; \quad V_{Hi} = V_H; \quad V_{Li} = V_L.$$

Через введенные параметры формулу (5.1) можно записать в виде:

$$-n1\sigma^{+}_{on}\left(V_g - V_L\right) - n2\sigma^{+}_{off}\left(V_g - V_L\right) + n3\sigma^{-}_{on}\left(V_H - V_g\right) + n4\sigma^{-}_{off}\left(V_H - V_g\right) + I_c = 0.$$

Из последнего равенства можно найти электрическое напряжение на шине

$$V_g = \frac{V_L\left(n1\sigma^{+}_{on} + n2\sigma^{+}_{off}\right) + V_H\left(n3\sigma^{-}_{on} + n4\sigma^{-}_{off}\right) + I_c}{\sigma_{full}}, \tag{5.2}$$

где $\sigma_{full} = n1\sigma^{+}_{on} + n2\sigma^{+}_{off} + n3\sigma^{-}_{on} + n4\sigma^{-}_{off}$ электрическая проводимость всех мемристорно-диодных связей, подключенных к шине.

Условием появления импульса на выходе схемы (рис. 2) является преодоление напряжения на шине некоторого порога срабатывания выходного инвертора. Порог срабатывания находится на уровне $k(V_H - V_L)$, где $k$ — безразмерный коэффициент, определяемый уровнем напряжения переключения по передаточной характеристике инвертора и имеющий значение около 0,5.

Так как логика срабатывания инверторов обратная, то указанный уровень на входе должен быть больше напряжения на шине $V_g$ для появления выходного импульса. Таким образом, прохождение импульса через электронный нейрон в реальной мемристорной 3D-матрице определяется условием:

$$k0{,}5\left(V \ -V_L\right) > \frac{V_L\left(n1\sigma^{+}_{on} + n2\sigma^{+}_{off}\right) + V_H\left(n3\sigma^{-}_{on} + n4\sigma^{-}_{off}\right) + I_c}{\sigma_{full}}. \qquad (5.3)$$

Условие (5.3) получено без учета динамических характеристик ячеек 3D логической матрицы. Необходимые характеристики могут быть легко введены в полную физическую модель, после снятия их параметров с лабораторных образцов. Также в представленном выражении можно учесть взаимовлияние соседних шин и возможные утечки между мемристорами по мемристивному материалу.

### 5.2.4. Анализ возможности получения максимального количества синаптических связей для суммации биоморфных импульсов

Биоморфная суммация импульсов выполняется в электронном нейроне на шине, к которой подключены мемристоры, играющие роль синаптических связей. Практический интерес представляет объединение достаточно большого количества мемристоров ($10^2$–$10^3$) на одной шине. Условие прохождения импульса (5.3) позволяет оценить количество электронных синаптических связей, которые рациональнее сделать на существующих мемристивных материалах. Основными параметрами формулы являются электрические проводимости мемристоров в крайних состояниях своего резистивного переключения. Также формулу (5.3) можно легко адаптировать для мемристоров, имеющих несколько устойчивых состояний электрической проводимости.

Источник тока $I_c$ для управления порогом срабатывания можно заменить на резистор с проводимостью $\sigma_c = I_c/(V_H - V_g)$. В этом случае, предполагая, что $V_L = 0$, формулу (5.4) можно переписать в виде:

$$V_g\left(n_1, n_2\right) = \frac{V_H\left(n_3\sigma^{-}_{on} + n_4\sigma^{-}_{off} + \sigma_A\right)}{n_1\sigma^{+}_{on} + n_2\sigma^{+}_{off} + n_3\sigma^{-}_{on} + n_4\sigma^{-}_{off} + \sigma_A}. \qquad (5.4)$$

Зависимость (5.4) для напряжения на шине представлена на графиках (рис. 5.7), построенных при двух значениях проводимости $\sigma_c$ для резисторного эквивалента источника тока. Значение проводимости мемристоров с учетом экспериментальных размеров и выбранного материала принято в интервале $10^{-3}$–$10^{-5}$ 1/Ом. Напряжение питания инвертора и значение $V_H$ равны 2000 мВ. Изначально предполагается, что в шину входят $10^5$ мемристорных связей всех режимов.

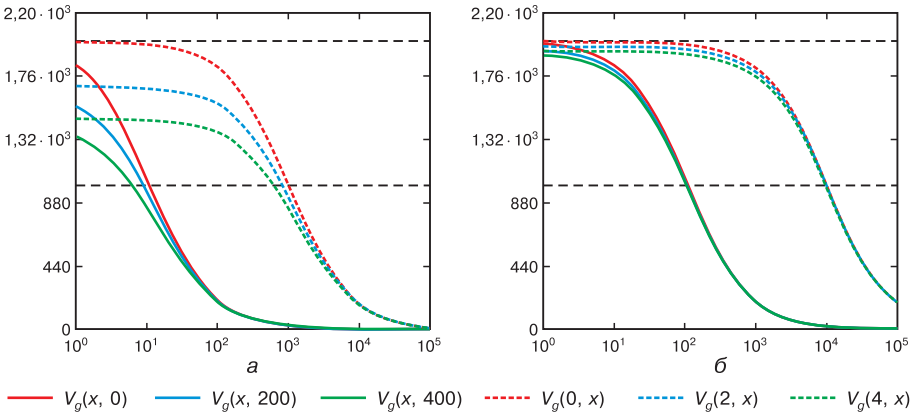



**Рис. 5.7.** Зависимости напряжения на шине от количества мемристоров первой и второй группы во время подачи входных импульсов:
*а* — графики построены для $\sigma_c = 10^{-2}$ 1/Ом; *б* — графики построены для $\sigma_c = 10^{-1}$ 1/Ом

Функция $V_g(n_1, n_2)$ на рис. 5.7 показана семейством кривых, полученных путем фиксирования одного из параметров. Мемристорные связи, работающие в режиме с обратносмещенным диодом Зенера (их количество обозначено $n_3$ и $n_4$), не оказывают на кривые заметного влияния из-за высокого электрического сопротивления. Сопротивление диода Зенера составляет $10^2$ и $10^9$ Ом в прямом и обратном направлениях соответственно. Поэтому увеличение количества двух оставшихся типов связей $n_1$ и $n_2$ проводилось за счет пренебрежимо малого изменения $n_3$ и $n_4$.

Можно видеть на графиках рис. 5.7, что с увеличением числа связей во время суммации информационных импульсов напряжение на шине понижается. В момент пересечения порога срабатывания, показанного в середине графиков пунктирной линией, на выходе инвертора будет получен импульс. Порог срабатывания инвертора задан на уровне $0{,}5V_H$. Условия прохождения информационных импульсов показывают кривые, обозначенные сплошными линиями. С увеличением числа высокопроводящих мемристорных связей, соответствующих переменной n1, условие прохождения импульса будет выполняться при превышении 10 и 100 связей в зависимости от $\sigma_c$, являющейся величиной проводимости эквивалента источника тока.

Возможно логически неправильное появление выходного импульса по причине утечки тока через мемристоры, находящиеся в низкопроводящем состоянии, при подаче на них входных информационных импульсов. Нижняя граница возникновения такой ситуации соответствует пересечению порога срабатывания кривыми на графиках, показанными пунктирными линиями на рис. 5.7. Ложный выходной импульс может появляться в результате суммирования чрезмерного количества входных импульсов, поступающих на «закрытые» мемристорные связи. Нежелательные срабатывания ограничены допустимым количеством входных импульсов. Эта величина определяется диапазоном резистивного переключения мемристорного материала.

На графике рис. 5.7, *а* видно значительное расхождение кривых, соответствующих проводимости $\sigma_c = 10^{-2}$ 1/Ом по сравнению с кривыми на рис. 5.7, *б* для проводимости $\sigma_c = 10^{-1}$ 1/Ом. Это обусловлено сильной зависимостью критерия срабатывания инвертора от количества поступающих импульсов, при малом токе источника, задающего порог срабатывания. Этот ток является варьируемой величиной для настройки режима работы каждой шины в 3D логической матрице.

Ток источника, задающего порог срабатывания, в данном случае определяется величиной проводимости резистивного эквивалента. Для порога срабатывания, равного $0{,}5V_H$, условие прохождения информационного импульса определяется неравенством:

$$\sigma_A(n_1, n_2) < n_1\sigma_{on}^{+} + n_2\sigma_{off}^{+}. \qquad (5.5)$$

По условию (5.5) на рис. 5.8 построены области, соответствующие режимам работы электронного нейрона в 3D логической матрице. Двухпараметрические зависимости показаны на графике путем фиксирования одного из параметров.

График на рис. 5.8, *а* показывает требуемую проводимость резисторного эквивалента тока, задающего порог срабатывания инвертора в режиме суммации импульсов в электронном нейроне. Интервал проводимости резистивного эквивалента тока электронного нейрона выбирается в областях «отсутствия импульсов» и «информационные импульсы» (рис. 5.8, *а*). Область «ложные импульсы» ограничивает количество связей для установленного порога срабатывания выходного инвертора.

График на рис. 5.8, *б* соответствует логическому режиму работы электронного нейрона. Логический режим работы отличается от режима суммации тем, что для генерации выходного импульса должно быть достаточно только одной низкопроводящей мемристорной связи, на которую поступает входной информационный импульс.

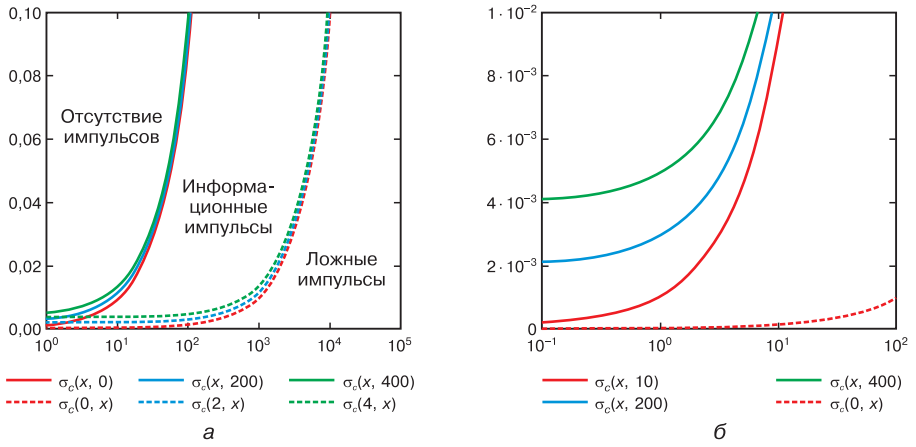


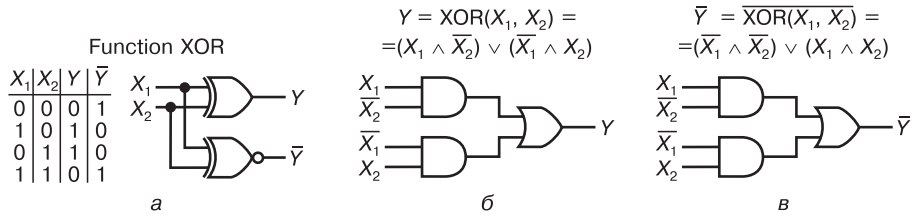

**Рис. 5.8.** Зависимости проводимости эквивалента источника тока для управления порогом срабатывания от количества мемристоров первой и второй группы в двух режимах работы электронного нейрона:
*а* — режим суммации импульсов; *б* — логический режим

### 5.2.5. Реализация логических функций на базе 3D КМОП-мемристорной логической матрицы

Во входном блоке нейропроцессора на этапе обработки данных, представленных в стандартном виде, необходимо выполнение базовых цифровых логических функций. Для реализации всевозможной логической функциональности в 3D КМОП-мемристорной логической матрице требуется программируемая коммутация логических элементов, обладающая функциональной полнотой. Под функциональной полнотой понимается возможность реализовать любые логические функции путем программирования мемристорных связей в 3D логической матрице.

В основу принципа работы входного блока нейропроцессора заложены программируемые дизъюнктивно нормальные формы (ДНФ). Для реализации ДНФ в 3D логической матрице достаточно сформировать полный логический базис. Архитектуре 3D логической матрицы возможна реализация логического базиса из операций И–НЕ и ИЛИ–НЕ при условии, что логические переменные будут подаваться в прямом и инверсном виде.

В качестве примера покажем реализацию логической функции исключающего ИЛИ (XOR). На рис. 5.9 показаны таблица истинности и схемы реализации с помощью базовых логических элементов И–НЕ и ИЛИ–НЕ функции XOR с прямым и инверсным выходами. Сверху изображения логических схем представлены совершенные ДНФ (СДНФ).




| $X_1$ | $X_2$ | $Y$ | $\overline{Y}$ |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 |

**Рис. 5.9.** Функции XOR с прямым и инверсным выходом:
*а* — таблица истинности; *б* — СДНФ для прямого выхода;
*в* — СДНФ для инверсного выхода

На входы схем поданы логические переменные $X_1$ и $X_2$ в прямом и инверсном видах. Как можно видеть, для получения прямого и инверсного состояний применяются одинаковые схемы с разной коммутацией входных логических переменных.

Применение логического базиса в 3D логической матрице показано на рис. 6 на примере реализации СДНФ логической функции XOR. Для реализации в 3D логической матрице любой СДНФ достаточно двух функциональных слоев. В слоях выполняется последовательно логическое И и инверсия с сигналами на шине. Логическая функция мемристорного переключателя определяется формулой:

$$\overline{\overline{X} \wedge M} \equiv X \vee \overline{M}.$$

В соответствии с законом де Моргана:

$$\overline{a \wedge b} = \overline{a} \vee \overline{b}$$

слои логической матрицы для каждого выхода выполняют функции по формуле:

$$\begin{aligned} Y_1 = &\left\{\left[\left(\overline{\overline{X_1} \wedge 1M_{11}}\right) \wedge \left(\overline{\overline{X_2} \wedge 1M_{21}}\right) \wedge \left(\overline{\overline{X_3} \wedge 1M_{31}}\right) \wedge \left(\overline{\overline{X_4} \wedge 1M_{41}}\right)\right] \wedge 2M_{11}\right\} \vee \\ \vee &\left\{\left[\left(\overline{\overline{X_1} \wedge 1M_{12}}\right) \wedge \left(\overline{\overline{X_2} \wedge 1M_{22}}\right) \wedge \left(\overline{\overline{X_3} \wedge 1M_{32}}\right) \wedge \left(\overline{\overline{X_4} \wedge 1M_{42}}\right)\right] \wedge 2M_{21}\right\} \vee \\ \vee &\left\{\left[\left(\overline{\overline{X_1} \wedge 1M_{13}}\right) \wedge \left(\overline{\overline{X_2} \wedge 1M_{23}}\right) \wedge \left(\overline{\overline{X_3} \wedge 1M_{33}}\right) \wedge \left(\overline{\overline{X_4} \wedge 1M_{43}}\right)\right] \wedge 2M_{31}\right\} \vee \\ \vee &\left\{\left[\left(\overline{\overline{X_1} \wedge 1M_{14}}\right) \wedge \left(\overline{\overline{X_2} \wedge 1M_{24}}\right) \wedge \left(\overline{\overline{X_3} \wedge 1M_{34}}\right) \wedge \left(\overline{\overline{X_4} \wedge 1M_{44}}\right)\right] \wedge 2M_{41}\right\}, \end{aligned}$$

что эквивалентно формуле

$$\begin{aligned} Y_1 = &\left\{\left[\left(X_1 \vee \overline{1M_{11}}\right) \wedge \left(X_2 \vee \overline{1M_{21}}\right) \wedge \left(X_3 \vee \overline{1M_{31}}\right) \wedge \left(X_4 \vee \overline{1M_{41}}\right)\right] \wedge 2M_{11}\right\} \vee \\ \vee &\left\{\left[\left(X_1 \vee \overline{1M_{12}}\right) \wedge \left(X_2 \vee \overline{1M_{22}}\right) \wedge \left(X_3 \vee \overline{1M_{32}}\right) \wedge \left(X_4 \vee \overline{1M_{42}}\right)\right] \wedge 2M_{21}\right\} \vee \\ \vee &\left\{\left[\left(X_1 \vee \overline{1M_{13}}\right) \wedge \left(X_2 \vee \overline{1M_{23}}\right) \wedge \left(X_3 \vee \overline{1M_{33}}\right) \wedge \left(X_4 \vee \overline{1M_{43}}\right)\right] \wedge 2M_{31}\right\} \vee \\ \vee &\left\{\left[\left(X_1 \vee \overline{1M_{14}}\right) \wedge \left(X_2 \vee \overline{1M_{24}}\right) \wedge \left(X_3 \vee \overline{1M_{34}}\right) \wedge \left(X_4 \vee \overline{1M_{44}}\right)\right] \wedge 2M_{41}\right\}. \end{aligned}$$

Для реализации функций XOR с прямым и инверсным выходами в соответствии с их СДНФ, представленной на рис. 5.9, мемристорные матрицы нужно определить значениями:

$$1M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} 2M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (5.6)$$

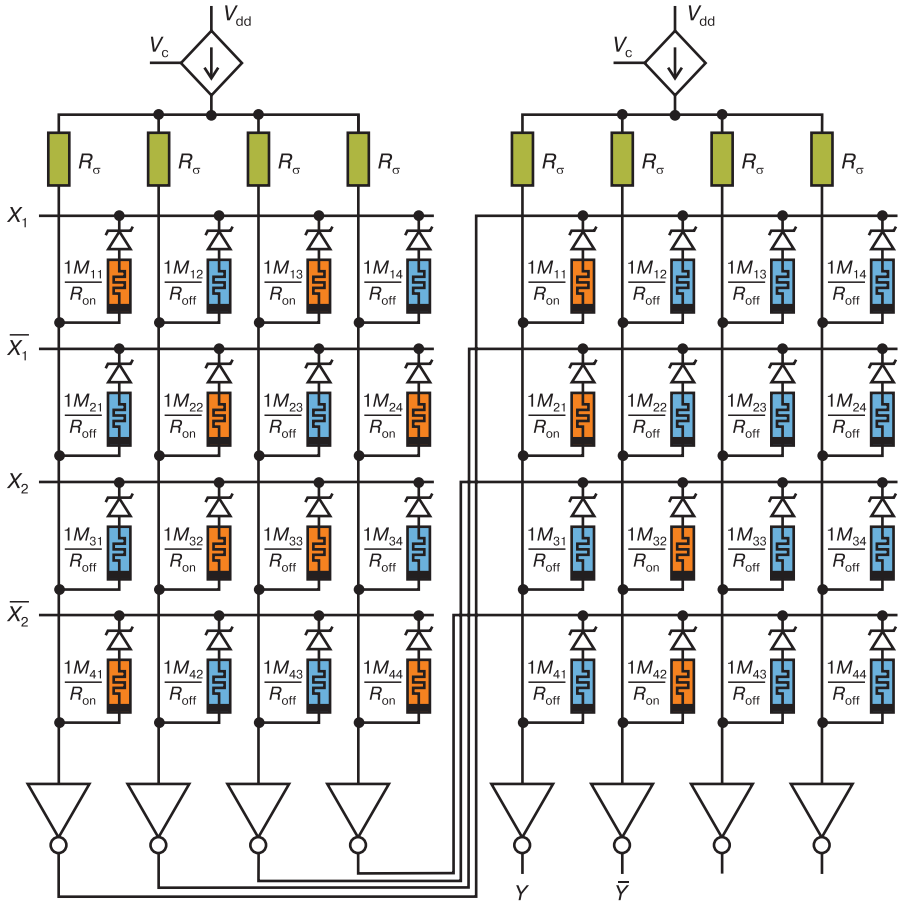



**Рис. 5.10.** Пример реализации логической функции исключающего ИЛИ в двух слоях 3D логической матрицы

На рис. 5.10 показана функция XOR, реализованная в соответствии с представленной формулой на базе двух слоев 3D логической матрицы. Мемристорные связи ($1M_{11}$–$1M_{44}$ и $2M_{11}$–$2M_{44}$), находящиеся

в низкопроводящем и высокопроводящем состояниях, отмечены на схеме синим и оранжевым цветами соответственно. Можно видеть, что «включенные» и «выключенные» мемристорные связи соответствуют расположению нулей и единиц в значениях $1M$ и $2M$ в равенстве (5.6).

Постоянные значения электрического тока, протекающего через сопротивления резисторов $R\sigma$, задают логический режим работы электронного нейрона.

Таким образом, показано, что 3D логическая матрица обладает функциональной полнотой, имея полный логический базис, необходимый для реализации комбинационных логических функций. На основе полного логического базиса, в том числе могут быть реализованы операции логического умножения и сложения данных, представленных в стандартном бинарном виде и поступающих во входной блок нейропроцессора. Информационные сигналы могут быть промодулированы биоморфными импульсами через источники тока, задающие режим работы электронного нейрона и установленные на периферии микросхемы.

### 5.2.6. SPICE-моделирование программирования резистивных состояний мемристора в 3D КМОП-мемристорной логической матрице

Мемриторы входного блока в 3D логической матрице требуют предварительного программирования в высокоомное или низкоомное состояние для использования их в реализации логических функций. Моделирование записи состояний мемристора $R_2$ показано на рис. 5.11. Для программирования используется инвертор, подключенный своим выходом к шине кроссбара, как показано на схеме рис. 5.11, *а*, который подачей высокого уровня напряжения на управляющий вход op/wr переведен в рабочий режим из высокоомного z-состояния. Вначале диаграммы (рис. 5.11, б) с помощью импульса ERASE на эпюре сигнала DL выполнено изменение сопротивления мемристора из низкоомного в высокоомное состояние для «стирания» соответствующей связи. После этого на диаграме показана подача импульса WRITE на эпюре сигнала $a_2$ и переход мемристора в низкоомное состояние для создания высокопроводящей связи.

Для моделирования изменений сопротивления мемристора была использована модель $R_2$, представленная в [25], с параметрами $R_{on} = 10kR_{off} = = 1000kR_{init} = 1000k$ beta $= 3e11V_t = 2{,}2$Cint $= 0{,}25p$.

Диоды Зенера $D_1$ и $D_2$, используются в схеме 3D логической матрицы как элемент для энергоэффективного выполнения логических функций. Принцип его работы в этом случае описан в работе [26]. Во время программирования состояний мемристоров диод Зенера находится в проводящем состоянии в обоих направлениях прохождения тока.

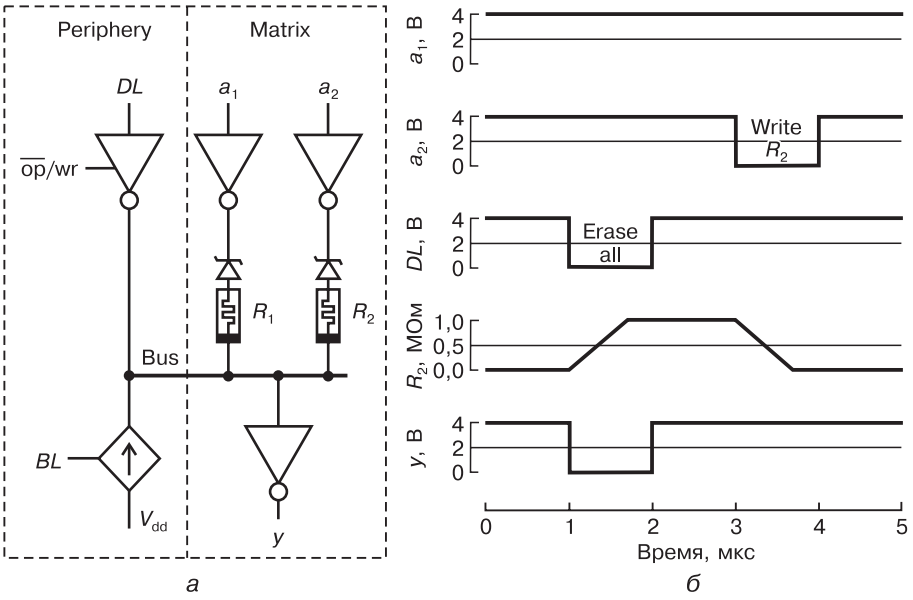



**Рис. 5.11.** Стирание и создание связи в 3D логической матрице:
*а* — схема моделирования;
*б* — эпюры сигналов и временные зависимости параметров схемы

Суть способа биоморфного кодирования информации заключается в том, что в качестве информационных сигналов используются сформированные картины импульсных последовательностей, аналогичных биологическим временным паттернам спайков. Такие импульсные последовательности формируются во входном блоке и передаются в центральные блоки нейропроцессора для последующей когнитивной обработки. Показано как с помощью импульсов — аналогов биологических спайков — осуществлять импульсное кодирование данных в электронном нейроне. Информационные параметры входного сигнала предложено кодировать частотой и пространственным распределением импульсов.

Для имитации этого механизма в электронной схеме предложено формирование регулируемых задержек распространения электрических импульсов, аналогичных задержкам спайков при их перемещении между нейронами. Промоделирована реализация электронных нейронов в 3D логической матрице при обработке импульсных последовательностей. Прохождение групп импульсов демонстрирует принцип суммации, характерный для работы биологического нейрона. Математически проанализированы условия реализации биоморфного импульсного способа кодирования информации в 3D логической матрице.

Во входном устройстве нейропроцессора спайки имитировались в виде коротких электрических импульсов. Биоморфное кодирование данных

осуществляется без учета сложных особенностей формы биологического потенциала действия. Это упрощение допустимо для входного блока нейропроцессора, поскольку в нем не требуется реализация функции обучения. Считается, что в биологических нейронных системах спайки являются основными носителями информации, а сложная форма потенциала действия ответственна за обучение, например, по механизму STDP.

Предложенный способ кодирования информации в электронном нейроне заслуживает особого внимания с точки зрения энергоэффективности передачи данных. Показано, что электронный нейрон затрачивает энергию только во время электрических импульсов. Тепловыделение Джоуля–Ленца происходит в основном на двух типах элементов схемы: на резисторах, играющих роль синапсов; на транзисторах схемы повторителя. Выделяется энергия во время действия импульса электрического тока, проходящего через резисторы. В повторителе выделяется тепло во время двух переключений транзисторов. Время импульса ограничено снизу скоростью срабатывания повторителя и наличием собственной электрической емкости шины.

Энергоэффективность нейропроцессора достигается за счет малой длительности информационных импульсов и минимизации количества импульсов, приходящихся на передачу единичного объема информации для предложенного биоморфного способа кодирования данных. Энергоэффективность дает возможность преодолеть тепловое ограничение масштабируемой технологии трехмерной компоновки элементов в мемристорных кроссбарах.

## 5.3. ИМПУЛЬСНОЕ СЖАТИЕ И КОДИРОВАНИЕ ЦИФРОВОЙ ИНФОРМАЦИИ ВО ВХОДНОМ УСТРОЙСТВЕ НЕЙРОПРОЦЕССОРА

### 5.3.1. Принцип работы входного блока нейропроцессора на основе логической матрицы

На вход входного устройства с интерфейсного блока поступает информационный сигнал в параллельном стандартном двоичном коде, в котором информация разделена на каналы по типу данных. Принцип работы входного блока: сначала происходит сжатие и нормализация входных данных с помощью дискретного косинусного преобразования в логической матрице, являющегося разновидностью методов Фурье-анализа, а затем амплитуды гармоник кодируются в биоморфные импульсы. Функциональная схема входного блока нейропроцессора на основе 3D логической матрицы представлена на рис. 5.12 [27].

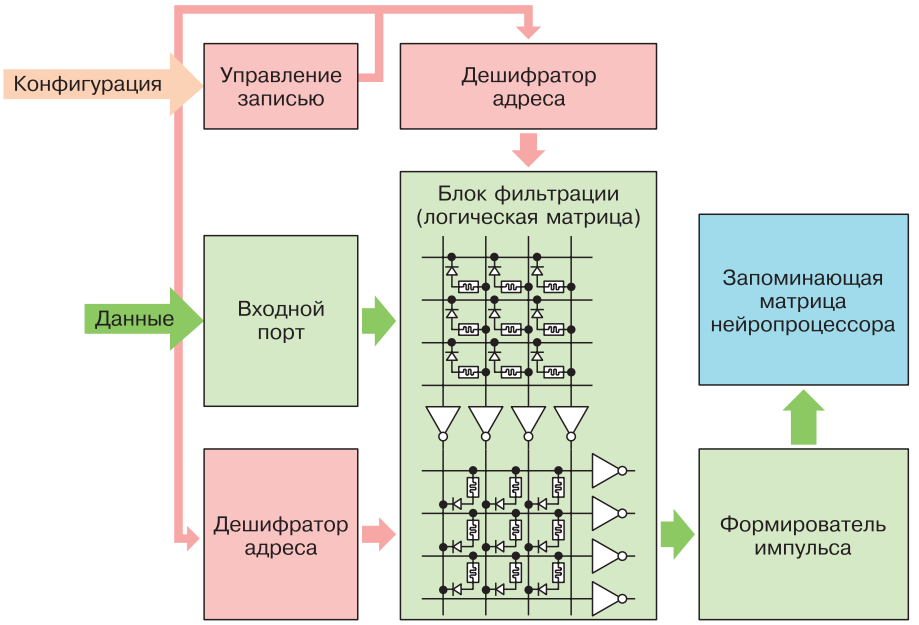



**Рис. 5.12.** Функциональная схема входного блока нейропроцессора на основе 3D логической матрицы

В основу принципа работы входного устройства нейропроцессора, сформулированного в предыдущем подразд. 5.2.5, заложены программируемые дизъюнктивно нормальные формы (ДНФ). Для реализации ДНФ в 3D логической матрице достаточно сформировать полный логический базис. Исходя из архитектуры логической матрицы, в ней возможна реализация логического базиса из операций И–НЕ и ИЛИ–НЕ при условии, что логические данные будут подаваться в прямом и инверсном видах. Подача тактовых импульсов на шины программирования и резисторов подтяжки инициирует импульсную работу электронного узла.

Пример применения полного логического базиса показан на рис. 5.13 в реализации совершенной ДНФ (СДНФ) логической операции исключающего ИЛИ. По такому же принципу реализованы операции логического умножения и сложения операндов, поступающих во входной блок нейропроцессора. Для реализации в логической матрице СДНФ достаточно двух функциональных слоев, при условии подачи в матрицу прямых и инверсных литералов каждой переменной.

Импульсная кодировка реализуется подачей управляющих сигналов на тактовые входы (C). Управляющие сигналы представляют собой импульсы, представленные на ВАХ коммутирующего элемента, в качестве которого выступает ячейка комбинированного кроссбара (рис. 5.14). В кроссбаре желтым цветом обозначены проводники, оранжевым — мемристорный слой, а фиолетовым и голубым — полупроводниковые слои диодов Зенера.

В основном логическом режиме работы (область I) 3D логической матрицы используются импульсы амплитудами $V_H$ и $V_L$, которые являются подпороговыми для мемристора, находящегося в составе коммутирующего элемента. В режиме смены логической функции входного устройства нейропроцессора на коммутирующий элемент подаются надпороговые импульсы $V_p$ и $-V_p$ (область II и III).

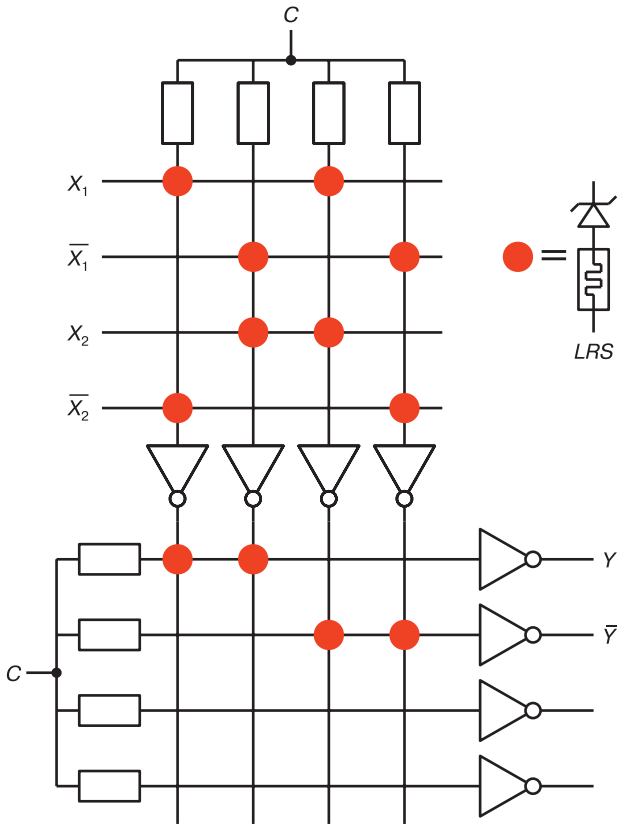



**Рис. 5.13.** Реализация исключающего ИЛИ в двух функциональных пластах логической матрицы

На рис. 5.14 показано, как значения яркости пикселей в строке видеокадра с помощью дискретного косинусного преобразования разлагаются в амплитуды гармоник. Гармоники с малой амплитудой фильтруются, а остальные с помощью перевода цифрового двоичного кода в пространственно-позиционный код преобразуются в формат биоморфных импульсов (рис. 5.15). В 3D логической матрице для каждого цифрового результата используется отдельная шина, на которую коммутируются импульсы исходных операндов через мемристорные переключатели.

В результате информационные импульсы появятся только на соответствующих входах запоминающей матрицы нейропроцессора, являющихся массивом синапсов нейронов. Низкие значения амплитуд отбрасываются, поэтому проводящих линий для них не предусмотрено.

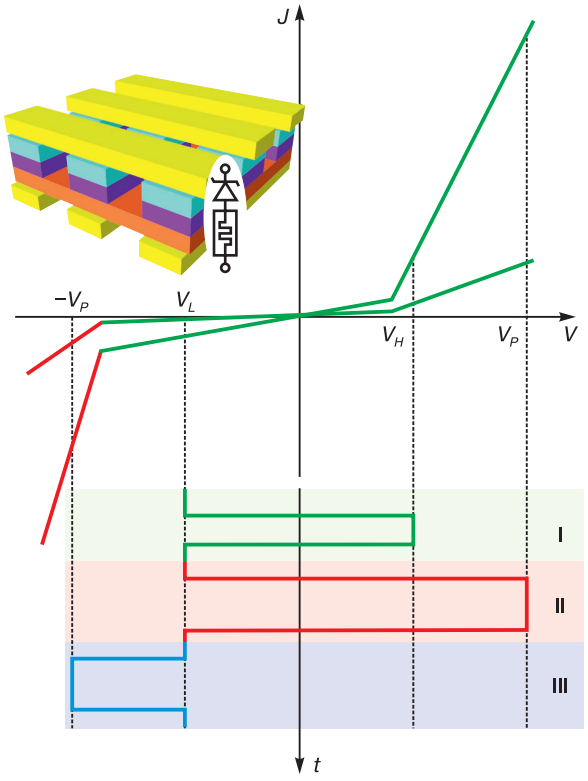



**Рис. 5.14.** Связь рабочих напряжений
с вольт-амперной характеристикой логической матрицы

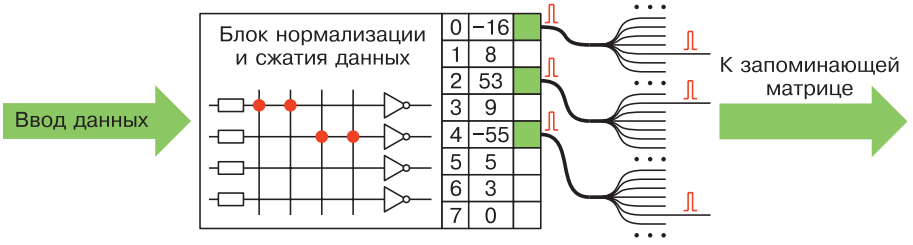



**Рис. 5.15.** Последовательность обработки цифровой информации
во входном блоке нейропроцессора: сжатие и кодирование значений
яркости пикселей в строке видеокадра в формат биоморфных импульсов

Таким образом, импульсное кодирование проводится в пласте 3D логической матрицы в два этапа. Сначала двоичный код преобразуется в высокий уровень напряжения на шине, используя базис в 3D логической матрице. После этого на шине формируется импульс или серия импульсов фиксированной длительности. Информация кодируется пространственным расположением шин и появлением на них электрических импульсов.

### 5.3.2. Генерация биоморфных импульсов

Формирование биоморфных импульсов с длительностью 2 мс обусловлено необходимостью создавать в мемристоре множество синаптических состояний для образования связей между нейронами в нейропроцессоре.

Преобразование информационных сигналов стандартного вида в биоморфный импульсный формат кодирования данных необходимо проводить во входном блоке нейропроцессора на этапе информационного преобразования. Для этого в схему, расположенную на периферии 3D логической матрицы, требуется ввести генератор биоморфных импульсов.

На рис. 5.16, *а* показана тестовая схема для моделирования работы 3D логической матрицы, содержащей на периферии генератор биоморфных импульсов ВР, который подключен к шине кроссбара с мемристорами через источник тока, управляемый напряжением (ИТУН). Мемристоры $R_3$ и $R_4$ запрограммированы в низкоомное состояние, мемристоры $R_1$ и $R_2$ имеют высокоомное состояние. Информационные электрические сигналы $x_1$–$x_4$, представленные в стандартном виде, через инверторы и шины кроссбара вышележащего слоя подаются на мемристоры $R_1$–$R_4$ соответственно. Конденсатор $C_1$ эмулирует емкость шины кроссбара. Инвертор линии декодера (DL) адреса относится к системе программирования, и на его выходе в этом режиме работы установлено высокоимпедансное Z-состояние путем подачи низкого уровня напряжения на управляющий вход op/wr.

На рис. 5.16, *б* представлены эпюры входных ($x_1$–$x_4$) и выходного ($y$) сигналов в двух режимах. Режим I соответствует подаче постоянного тока от ИТУН на шину кроссбара. В режиме II реализована импульстная подача тока от ИТУН, управляемым биоморфными импульсами по линии ВР. Диаграмма показывает, что в представленной схеме реализуется логика ИЛИ для сигналов $x_3$ и $x_4$. Влияние сигналов $x_1$ и $x_2$ на шину кроссбара подавлено высокоомными состояниями мемристоров $R_1$ и $R_2$ до уровня ниже порога переключения выходного инвертора. В режиме II в отличие от режима I на эпюре выходного сигнала ($y$) широкий импульс заполнен сигналом генератора биоморфных импульсов.

Предложенный способ ввода импульсов в 3D логическую матрицу входного блока нейропроцессора, также позволяет модулировать пространственное кодирование информации импульсными пачками.

Сформированные пачки импульсов могут быть использованы для перепрограммирования мемристоров нейропроцессора в процессе имитирования синаптической пластичности принципа STDP (*spike-timing-dependent plasticity*) и ассоциативного самообучения-разобучения, принципы которого представлены в [28].

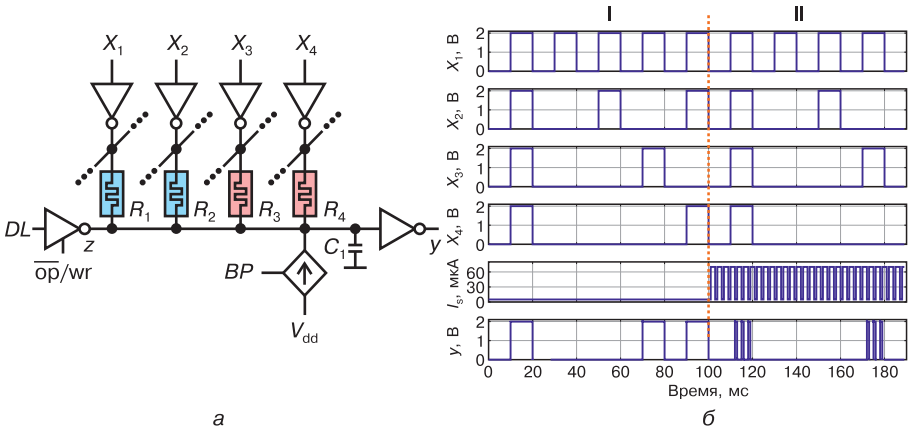



**Рис. 5.16.** Генерация биоморфных импульсов во входном блоке нейропроцессора:

*а* — схема моделирования; *б* — эпюры сигналов

## 5.4. СПОСОБЫ КОДИРОВАНИЯ ИНФОРМАЦИИ В ИМПУЛЬСЫ

Существует несколько способов кодирования информации, представляющей собой набор величин (яркости пикселей, амплитуды составляющих частот звукового сигнала) с определенными значениями, в импульсы: кодирование одной величины на индивидуальном входном нейроне, позиционное кодирование и промежуточный вариант — популяционное кодирование (рис. 5.17).

Позиционное кодирование отображает входные значения по принципу «значение–позиция», т. е. каждому уникальному значению соответствует свой входной нейрон. Очевидно, что такой способ кодирования является неэффективным из-за использования большого количества нейронов.

Один входной нейрон и небольшая популяция входных нейронов могут кодировать величину в частоту [29] или в задержку импульсов [29–31]. Кодирование в задержки потенциально позволяет передать значение величины за один импульс [32]. При этом осуществляется выборка значений через интервал времени, превышающий максимально возможную задержку. Такая выборка приводит к потере информации.

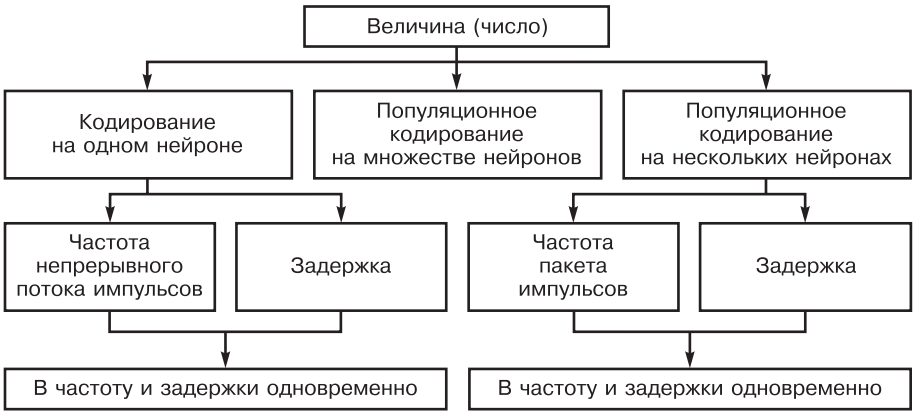



**Рис. 5.17.** Способы кодирования величин в зависимости от используемого числа входных нейронов

Кодирование в частоту может осуществляется как для пакета импульсов, так и для непрерывной череды импульсов [33]. Этот способ кодирования имеет недостаток, связанный с потерей времени и заключающийся в том, чтобы определить частоту следования, необходимо принять несколько импульсов.

Возможность аппаратной реализации одновременного (комбинированного) кодирования информации на одном входном нейроне в частоту и задержки представлена в [34]. В частоту кодируется само значение кодируемой величины, а в задержку — производная этой величины во времени, что позволяет нейросети быстрее реагировать на изменения входного сигнала за счет большего объема передаваемой информации в единицу времени.

Популяционное кодирование отличается от кодирования на одном входном нейроне тем, что несколько нейронов позволяют быстрее обрабатывать величину, имеющую множество значений. Это кодирование позволяет обеспечить избыточность информации с наименьшими накладными расходами по сравнению с другими способами кодирования за счет перекрытия рецепторных полей (настроечных кривых) входных нейронов [31]. Избыточность представления информации увеличивает отказоустойчивость способа кодирования. С другой стороны, для реализации такого кодирования требуется электрическая схема с большим количеством элементов. Способ популяционного кодирования, наблюдаемый в биологических нейронных сетях [12], учитывает пространственную производную входного числа (яркости) в отличие от учета временной производной в [34]. Использование популяционного кодирования информации, обеспечивающего высокую отказоустойчивость и больший объем передаваемой информации, перспективно во входном устройстве биоморфного нейропроцессора. В настоящее время частично реализовано аппаратное

популяционное кодирование в частоту на транзисторных СБИС [35] и в задержки с помощью ПЛИС [36]. Для построения законченной схемы кодирующего устройства в эти схемы необходимо добавить преобразователь выходного сигнала в импульсы. Использование указанных устройств во входном блоке биоморфного нейропроцессора не эффективно с точки зрения занимаемой площади и энергопотребления. Далее приводится универсальная электрическая схема с использованием комбинированного мемристорно-диодного кроссбара, способная выполнять разные виды кодирования, и исследуется ее работа.

## 5.5. АППАРАТНОЕ КОДИРОВАНИЕ ИНФОРМАЦИИ В ИМПУЛЬСЫ НА ОСНОВЕ МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА

В подразд. 5.3.1 представлена схема кодирования на основе логической матрицы с комбинированным мемристорно-диодным кроссбаром [1; 2], в которой полученные в результате сжатия гармоники с помощью перевода цифрового двоичного кода в пространственно-позиционный код преобразуются в формат биоморфных импульсов.

Рассмотрим схему кодирования информации на основе двух логических матриц [2] и линий задержек, в которой преобразование выполняется в матрице конъюнкций: входной двоичный код разрешает прохождение определенных импульсов от линий задержек в матрицу дизъюнкций. Схема является универсальной, поскольку позволяет реализовать следующие способы импульсного кодирования входной информации, представленной двоичными числами: кодирование в частоту и в задержки на одном виртуальном входном нейроне и на их популяции, а также одновременное кодирование в частоту и задержки популяцией нейронов.

Функциональная схема кодирующего устройства для одного числа представлена на рис. 5.18. При кодировании нескольких чисел (яркостей пикселей, амплитуды образа косинусного преобразования) необходимо использовать несколько кодирующих устройств параллельно.

Кодирующее устройство одного числа содержит две программируемые логические матрицы [2]. Электрическая схема кодирующего устройства, реализующая все указанные виды импульсного кодирования, представлена на рис. 5.19. Для простоты схема на рисунке имеет двухбитный вход, две линии задержки и два выходных канала.

В мемристорно-диодном кроссбаре логической матрицы реализуются логические вентили И (конъюнкции) на основе диодно-резистивной логики с возможностью отключения любых входов вентиля путем изменения сопротивления мемристоров. Инверторы на выходе служат для восстановления значений напряжений логических уровней.

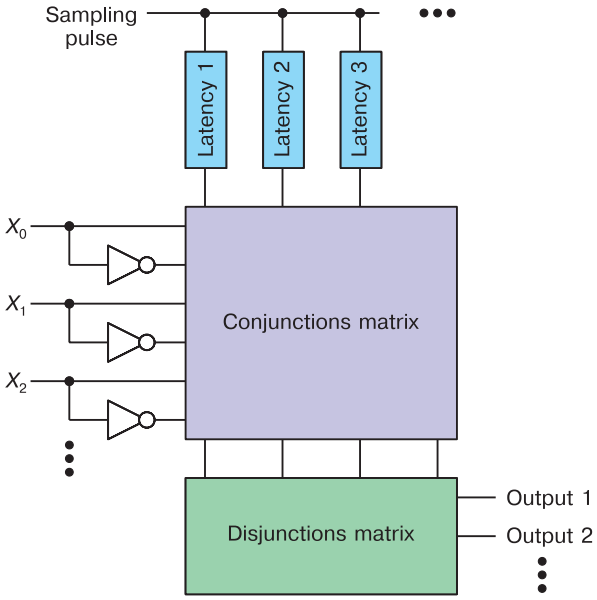



**Рис. 5.18.** Функциональная схема кодирующего устройства одного числа

В первой матрице вместо источника напряжения для подтягивающих резисторов подключены линии задержки, а вторая матрица используется без изменений. Матрицы подключены последовательно и запрограммированы на реализацию совершенной дизъюнктивной нормальной формы, коммутирующей импульсы с линий задержек на выходы в зависимости от входного двоичного числа. Точность представления входного числа определяется количеством используемых бит, а точность выходного отображения — от набора величин задержек. Кодирование происходит при поступлении от управляющей схемы импульса, запускающего выборку входного числа. Линии задержки построены на базе двух *RC*-цепей и логических вентилей. Постоянная времени первой интегрирующей *RC*-цепи определяет величину задержки. Задержка является программируемой величиной, поскольку в качестве резистора использован мемристор. Вторая *RC*-цепь отвечает за ширину выходного импульса.

Для увеличения разрядности входного числа необходимо добавить соответствующие горизонтальные проводники в первой логической матрице. Увеличение числа доступных задержек приведет к увеличению числа вертикальных проводников в обеих матрицах. Число выходов соответствует числу горизонтальных проводников мемристорно-диодного кроссбара второй логической матрицы.

Расчет кодирующего устройства выполнялся в оригинальной специализированной программе MDC–SPICE для расчета больших электрических схем, содержащих мемристорно-диодные кроссбары. Этот симулятор является модифицированной версией SPICE подобного симулятора NGSPICE.

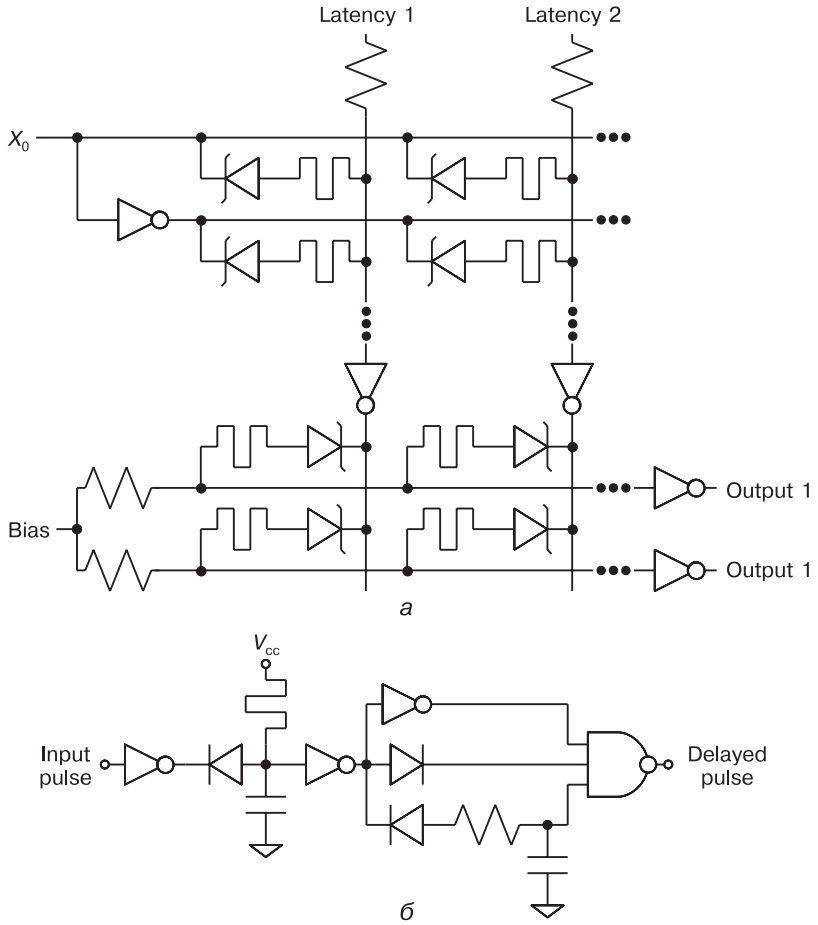



**Рис. 5.19.** Электрическая схема кодирующего устройства:
*а* — реализация совершенной дизъюнктивной нормальной формы на базе мемристорно-диодного кроссбара; *б* — линия задержки

При расчете больших схем используется модель мемристора [25], в которой изменения параметра состояния дополнительно были жестко зафиксированы в интервале от 0 до 1. Такое ограничение необходимо, поскольку неабсолютная точность рациональных чисел в компьютерной системе приводит к выходу параметра состояния за границы допустимого интервала и, как следствие, к неправильной работе модели. Кроме этого, с целью ускорения расчета нелинейная вольт-амперная характеристика диода Зенера была упрощена и представлена в виде трех прямых линий. То есть симулятор заменяет диод резистором с соответствующим значением сопротивления. Вносимая при упрощении ошибка мала, когда мемристорно-диодный кроссбар работает в цифровом режиме, и напряжение на диодах Зенера не приближается к пороговым значениям открытия и обратного пробоя диода.

### 5.5.1. Кодирование числа в частоту импульсов

Кодирование в частоту означает, что частота выходных импульсов схемы пропорциональна входному двоичному числу. Этого можно добиться, запрограммировав линии задержки в электрической схеме кодирующего устройства (см. рис. 5.3) таким образом, чтобы задержки были кратны некоторому временному интервалу, который должен быть больше ширины импульсов. Логическая схема разрешает или запрещает прохождение импульсов от линии задержек на выход в зависимости от текущих значений битов входного числа. Коэффициент пропорциональности при преобразовании можно изменять, перенастраивая логические матрицы на использование разных линий задержки. Количество задержанных импульсов, прошедших на выход, будет определять среднюю частоту выходных импульсов за время выборки.

#### *5.5.1.1. Один виртуальный входной нейрон*

Если производится кодирование одного входного числа одним входным виртуальным нейроном, то выход в кодирующем устройстве (см. рис. 5.18) будет один и, соответственно, реализована одна дизъюнкция во второй логической матрице (см. рис. 5.19). Число задействованных входов будет равно числу бит входного числа.

Для демонстрации работы кодирующего устройства в этом режиме на вход подавалась линейно возрастающая во времени последовательность двоичных чисел от нуля до максимального значения (рис. 5.20, *а*). Полученные в ходе SPICE-моделирования соответствующие выходные импульсы приведены на рис. 5.20, *б*.

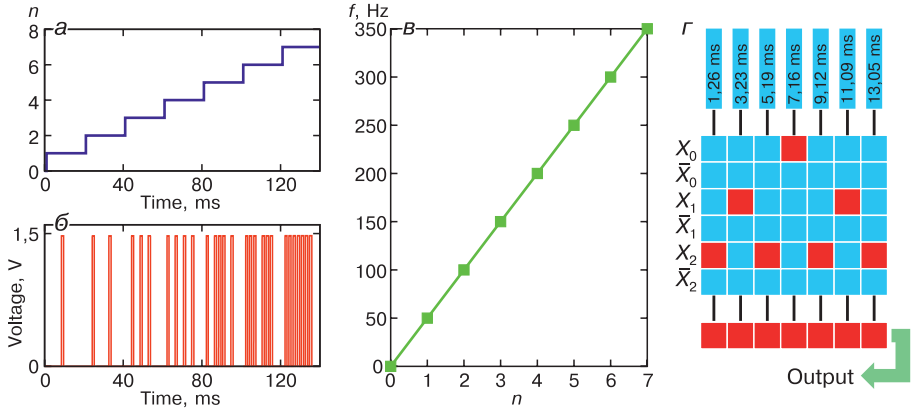



**Рис. 5.20.** Результат SPICE-моделирования режима кодирования числа в частоту на одном нейроне:

*а* — изменение входного числа во времени; *б* — выходные импульсы; *в* — функция преобразования числа *n* в частоту *f*; *г* — карта проводимости мемристоров

Из рис. 5.20, *а* и рис. 5.20, *б* видно, что в результате процедуры кодирования каждому числу *n* соответствует пакет импульсов со средней частотой *f*, а монотонно возрастающая функция преобразования (рис. 5.20, *в*) обеспечивает однозначность преобразования. Энергопотребление схемы в среднем составило 39,2 мкВт.

### 5.5.1.2. Популяция виртуальных входных нейронов

Если кодирование одного числа производится популяцией нейронов, то будет задействовано несколько выходов в матрице дизъюнкций (см. рис. 5.18, 5.19). При этом импульсы, которые ранее передавались по одному каналу, теперь принудительно распределяются на выходные каналы кодирующего устройства, причем суммарная частота пропорциональна кодируемому числу.

На рис. 5.21 приведены результаты SPICE-моделирования процесса кодирования числа в частоту схемой (см. рис. 5.19), запрограммированной в режим работы трех виртуальных нейронов (выходных каналов).

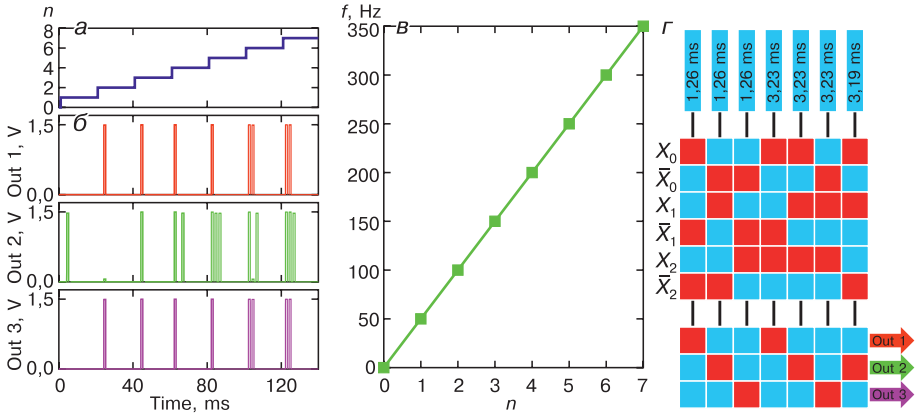



**Рис. 5.21.** SPICE-моделирование режима кодирования популяцией из трех нейронов:

*а* — изменение входного числа во времени; *б* — выходные импульсы; *в* — функция преобразования числа *n* в частоту *f*; *г* — карта проводимости мемристоров

Из рис. 5.21 следует, что суммарная частота импульсов на выходах пропорциональна входному числу. Поскольку импульсы перераспределены по нескольким каналам, время передачи значения входного числа сокращается. Так, на рис. 5.20 максимальное время передачи одним виртуальным нейроном составляет 13 мс, а тремя виртуальными нейронами на рис. 5.5 — 5 мс. То есть скорость передачи возросла в 2,6 раза. Однако кодирование входного числа в частоту импульсов популяцией нейронов требует большего числа строк в матрице дизъюнкций и, соответственно, добавлено 14 ячеек 1D1M в ее кроссбаре и два инвертора. Причем энергопотребление увеличилось незначительно — на 84,9 мкВт.

### 5.5.2. Кодирование числа в задержки импульсов

При этом кодировании входное двоичное число преобразуется в задержку информационных импульсов относительно времени начала выборки. Величина задержки определяется так называемой tuning curve (настроечная кривая) нейрона, в качестве которой частоты выбирают функцию Гаусса без нормирующего множителя [31; 37]. Настроечная кривая нейрона обеспечивает тем меньшую задержку импульса, чем ближе кодируемая величина к собственному значению данного нейрона. Нужные величины задержек в электрической схеме устройства (см. рис. 5.19) можно запрограммировать, изменяя сопротивления мемристоров в линиях задержки. Первая логическая матрица запрограммирована на работу как мультиплексор. Она пропускает через себя импульс с линии задержки, отвечающей текущему входному числу. Вторая логическая матрица собирает эти прошедшие импульсы на выходы устройства.

#### *5.5.2.1. Один виртуальный входной нейрон*

Матрица дизъюнкций и количество выходов при кодировании одного входного числа в задержку импульса виртуальным нейроном будут такими же, как при частотном кодировании. В отличие от частотного кодирования, линии задержки в электрической схеме кодирующего устройства (см. рис. 5.19) перенастроены с учетом настроечной кривой нейронов (например, гауссианов). Соответственно, матрица конъюнкций пропускает через себя импульсы с задержкой, согласно настроечной кривой нейрона.

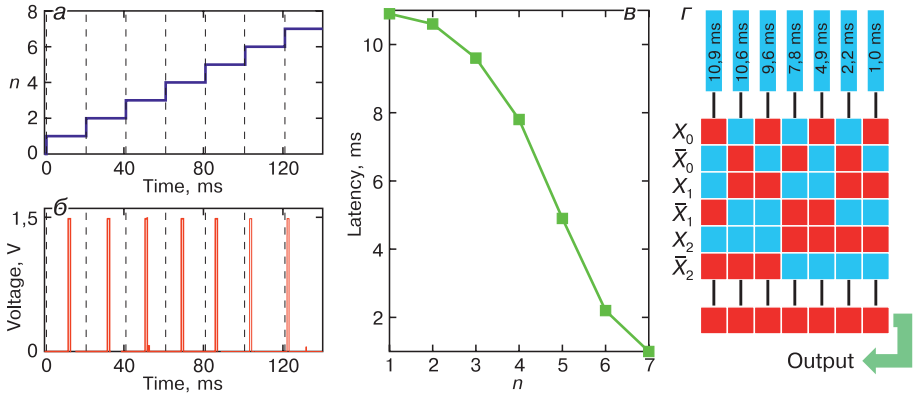



**Рис. 5.22.** SPICE-моделирование режима кодирования числа в задержку импульса:

*а* — изменение входного числа во времени; *б* — выходные импульсы; *в* — функция преобразования числа *n* в задержки; *г* — карта проводимости мемристоров в блок-схеме устройства

При SPICE-моделировании работы кодирующего устройства в этом режиме на вход подавалась такая же линейно возрастающая во времени последовательность двоичных чисел от нуля до максимального значения (рис. 5.22, *а*), как и при частотном кодировании. На рис. 5.22, *б* показаны выходные импульсы, полученные в ходе моделирования.

Из рис. 5.22, *а* и *б* видно, что в результате процедуры кодирования каждому числу $n$ соответствует импульс с задержкой $\tau$ относительно импульса выборки значения входного числа. На рис. 5.22, *в* показана получившаяся функция преобразования, которая соответствует использованному при кодировании гауссиану преобразования. Для обеспечения этого режима кодирования мемристоры в логических матрицах кодирующего устройства были запрограммированы согласно рис. 5.22, *г*. Красный цвет соответствует низкому сопротивлению мемристора, а синий — высокому. Энергопотребление схемы в среднем составило 54 мкВт.

### 5.5.2.2. Популяция виртуальных входных нейронов

Матрица дизъюнкций (см. рис. 5.18, 5.19) в этом режиме кодирования имеет несколько выходов, соответствующих популяции виртуальных входных нейронов. Каждый нейрон имеет собственное значение (максимум гауссиана), при котором его задержка минимальна. Исходя из настроечных кривых, при кодировании любого входного числа каждый нейрон выдаст по одному импульсу. Поскольку областью значений гауссианов является бесконечный интервал, максимальная задержка импульсов должна быть ограничена.

На рис. 5.23 приведены результаты SPICE-моделирования процесса популяционного кодирования числа в задержки импульсов схемой, представленной на рис. 5.3, включающей 3 нейрона.

Максимальные отклики нейронов соответствовали входным числам 1, 4 и 7. На вход подавалась линейно возрастающая во времени последовательность трехразрядных двоичных чисел от нуля до максимального значения, а затем резкое изменение до 1 и возврат до 7 (рис. 5.23, *а*).

При сравнении с рис. 5.22, демонстрирующим работу кодирующего устройства с виртуальным одним нейроном, выходные сигналы на рис. 5.23, *б* имеют другую задержку, поскольку каждый виртуальный нейрон имеет свою настроечную кривую (рис. 5.23, *в*) в виде гауссиана с меньшим среднеквадратичным отклонением по сравнению с единственным нейроном. Кроме этого, мгновенные частоты нейронов отражают скорость изменения значения входного числа за выборку: при изменении с 7 на 1 выходная частота первого нейрона возрастет, поскольку задержка изменится от максимальной до минимальной. На третьем выходе реализуется тот же механизм при изменении входного числа с 1 на 7. Поведение нейронов противоположно, потому что они имеют зеркально отраженные половинки гауссианов. Фактически в частоты нейронов будет закодирована временная производная входного числа, так же как в [34] для одного нейрона.

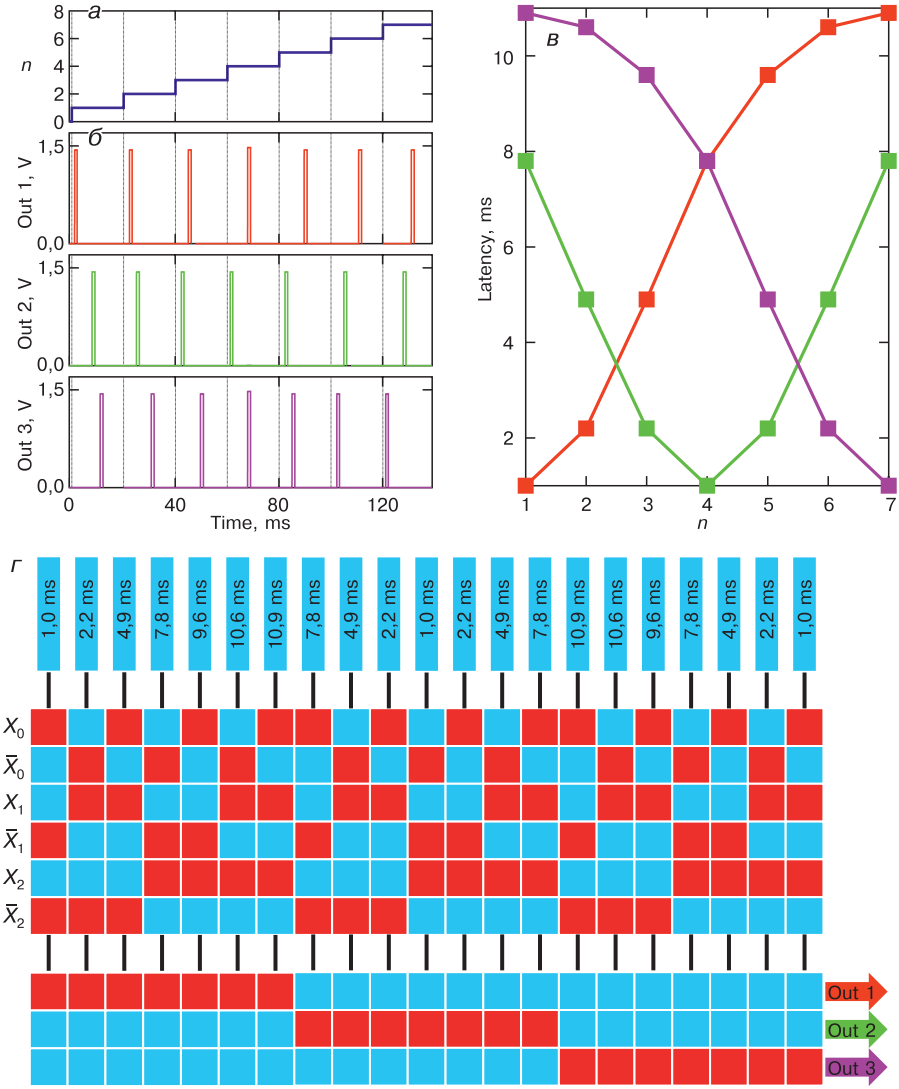



**Рис. 5.23.** SPICE-моделирование режима кодирования в задержки импульсов популяцией из трех нейронов:

*а* — изменение входного числа во времени; *б* — импульсы на выходных каналах; *в* — функции преобразования числа *n* в задержки для каждого канала; *г* — карта проводимости мемристоров в блок-схеме устройства

На рис. 5.23, *г* отражено состояние запрограммированных мемристоров в схеме с тремя выходными каналами. Соответственно, в матрице дизъюнкций задействовано три строки. А матрица конъюнкций содержит в три раза больше столбцов для обеспечения работы линий задержек для трех настроечных кривых. По сравнению со схемой кодирования в задержки

одним нейроном добавлено 84 ячейки кроссбара и 14 инверторов в матрице конъюнкций и 63 ячейки кроссбара и 2 инвертора в матрице дизъюнкций. Энергопотребление при этом возросло в 3,8 раза и составило 205 мкВт.

Таким образом, достигнута избыточность представления входного числа за счет передачи нескольких (в данном случае трех) независимых информационных импульсов по отдельным каналам. Даже если один из каналов окажется нерабочим, то информация будет передана с меньшей точностью, за счет настроечных кривых соседних каналов, перекрывающих настроечную кривую нерабочего канала. Это увеличивает отказоустойчивость способа кодирования.

Помимо повышенной отказоустойчивости схема с тремя нейронами позволяет увеличить скорость передачи информации за счет уменьшения высоты гауссианов преобразования и соответствующего сокращения числа линий задержек при кодировании. При уменьшении высоты гауссианов линейно возрастает скорость кодирования и линейно уменьшается максимально возможное число задержек. Минимальная высота гауссиана достигается при использовании двух задержек.

### 5.5.3. Одновременное кодирование популяцией нейронов пространственной производной входного числа в частоту и значения входного числа в задержки импульсов

Функциональная схема одновременного кодирования и ее соответствующая электрическая схема построены путем параллельного включения нескольких схем популяционного кодирования в задержки импульсов (см. рис. 5.23). При этом импульсы с одинаковых выходов перенаправляются на общий выход логическим элементом ИЛИ. Если входные числа отдельных схем кодирования, например, отвечающие яркости соседних пикселей изображения, отличаются, то выходные частоты нейронов увеличатся, поскольку соответствующие задержки различаются. Кроме производных входных чисел по времени в этом случае в частоту будет закодирована и производная яркости пикселя по пространственным координатам изображения.

На рис. 5.24 представлен результат SPICE моделирования процесса кодирования двух входных чисел $n_0$ и $n_1$ в задержки трех виртуальных нейронов. Число $n_0$ линейно возрастало во времени от 1 до 7, а $n_1$ наоборот убывало с 7 до 1 (рис. 5.24, *а*). Настроечные кривые нейронов (рис. 5.24, *в*) такие же как при моделировании кодирования одного числа в задержку популяцией нейронов.

Входные числа $n_0$ и $n_1$ кодируются своими подсхемами в задержки импульсов. Поскольку выходы двух кодирующих подсхем объединены логикой ИЛИ, выходные сигналы смешиваются. Если входные числа $n_0$ и $n_1$ равны, то на выходе присутствует только по одному импульсу на каждом канале.

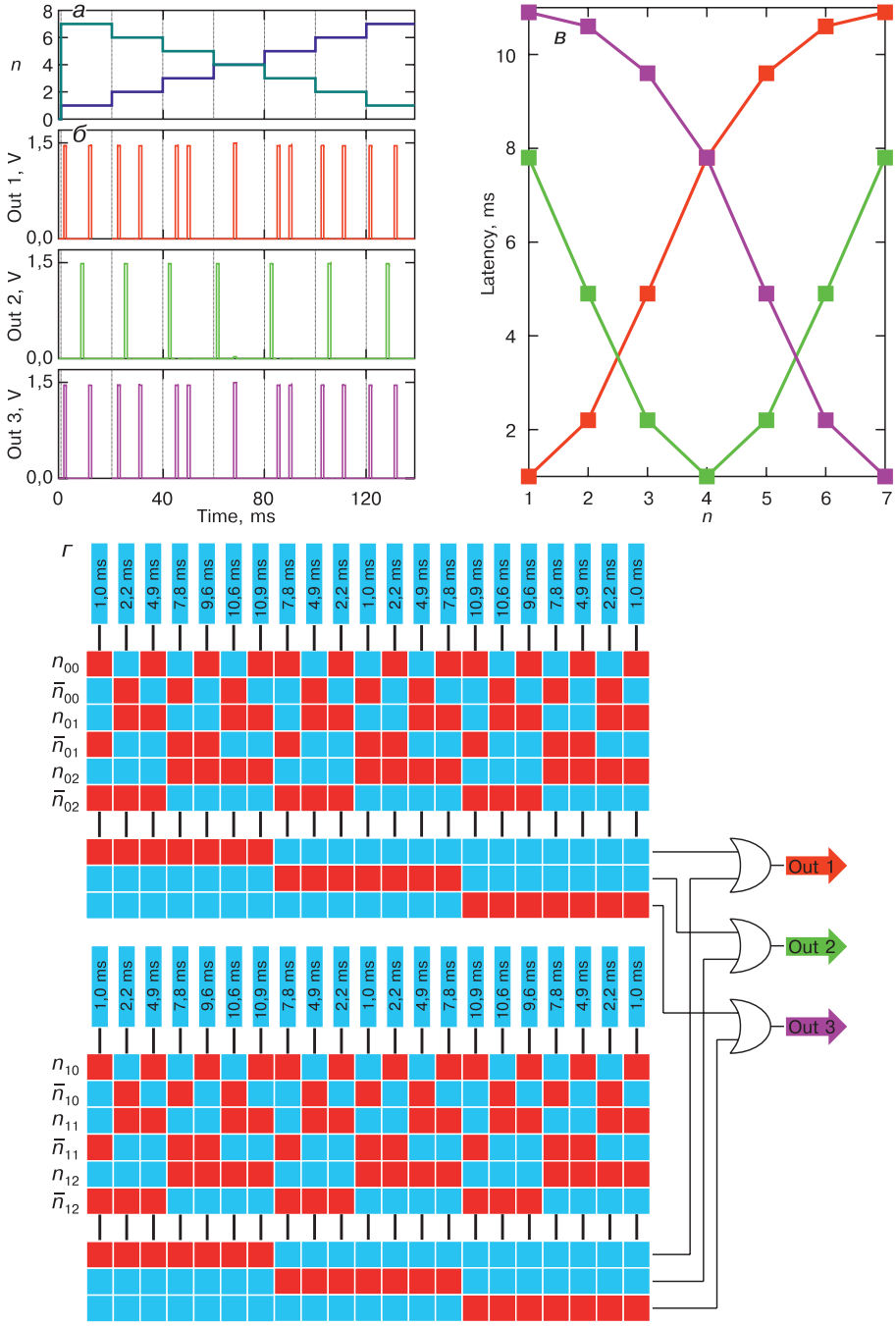



**Рис. 5.24.** SPICE-моделирование режима кодирования одновременно в задержки и в частоту импульсов популяцией нейронов

Если же $n_0$ и $n_1$ различаются, то на каждом выходном канале наблюдаются два импульса. Таким образом, при различии в значениях $n_0$ и $n_1$ выходная частота удваивается, что соответствует отличию от нуля пространственной производной входных чисел (яркости пикселей).

Аппаратная реализация функции преобразования для популяционного кодирования на транзисторных ПЛИС [22] требует 102,2 тысячи логических вентилей, состоящих минимум из двух транзисторов, на один канал. Для обслуживания 500 входов запоминающей матрицы потребуется ПЛИС с более 50 млн вентилей. В то время как при использовании мемристорно-диодного кроссбара число вентилей равно удвоенному числу линий задержек. Для достижения такой же точности, что и при использовании 8-битных чисел в ПЛИС, предлагаемая схема на мемристорном кроссбаре с использованием смешанных аналогово-цифровых вычислений потребует 256 линий задержек, и соответственно, число инверторов в одном выходном канале будет равно 512. В ПЛИС [22], в которой реализованы 24 входных нейрона, максимальная потребляемая мощность составляет 2 Вт, то есть 83 мВт на канал. В то время как в кроссбаре один канал потребляет 54 мкВт. СБИС-реализация функции преобразования [21] при точности чисел 8 бит требует 3344 транзистора на канал. Поскольку эта схема является полностью аналоговой, она потенциально имеет высокое энергопотребление. Для построения полноценного кодирующего устройства необходима схема преобразования выходного сигнала в импульсы, которая потребует дополнительной площади на кристалле. Например, в ПЛИС такой преобразователь можно реализовать на основе счетчика и логического компаратора, а для схемы [21] — на основе интегратора и триггера Шмитта. Таким образом, применение логических матриц на основе мемристорно-диодного кроссбара в кодирующем устройстве позволяет минимизировать занимаемую площадь на кристалле и энергопотребление устройства.

Одновременное кодирование значения яркости пикселей в задержки и пространственной производной яркости в частоту наблюдается в биологических нейронных сетях и позволяет передать больше визуальной информации за то же время [20]. Предложенная схема одновременного кодирования популяцией нейронов пространственной производной входного числа в частоту и значения входного числа в задержки импульсов дополнительно позволяет кодировать и производную входной величины во времени. Такая схема более предпочтительна по сравнению с возможной аппаратной реализацией [19] одновременного кодирования на одном нейроне значения входной величины в частоту и в задержку ее производной во времени. Она позволяет быстрее реагировать на изменения входного сигнала за счет большего объема передаваемой информации.

Режим одновременного кодирования популяцией виртуальных нейронов передает большее количество информации по сравнению с другими режимами, поскольку учитывает значение кодируемого числа вместе с его производными в пространстве и во времени. Подобное кодирование наблюдается в биологических нейронных сетях [20], но без учета производной во времени.

Кодирующее устройство способно формировать импульсы напряжения произвольной длительности, в том числе биоморфные импульсы длительностью 1 мс. Искусственные нейронные сети с такими импульсами будут работать со скоростью биологических нейронных сетей, что можно применить в биотехнологиях, например, в нейропротезировании.

Использование логической матрицы на основе мемристорно-диодного кроссбара как в основных частях нейропроцессора, так и во входном устройстве, позволяет унифицировать элементную базу и источники питания нейропроцессора.

## 5.6. ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ ОБ АКТИВАЦИИ НЕЙРОНОВ В ЦИФРОВОЙ ДВОИЧНЫЙ КОД В ВЫХОДНОМ УСТРОЙСТВЕ НЕЙРОПРОЦЕССОРА

### 5.6.1. Функциональная характеристика выходного устройства

Основной задачей выходного устройства является преобразование информации из импульсного формата в стандартный цифровой код для вывода из нейропроцессора. В выходном устройстве происходит сбор групп распределенных сигналов аппаратной нейросети нейропроцессора после их параллельной обработки. Эти функции необходимы, поскольку данные в нейропроцессоре в формате своего представления могут охватывать большую группу сигнальных линий. Осуществляется представление обработанных величин в сжатом формате (без избыточности), локализованном по времени и пространству.

На выходное устройство поступает информация, полученная в нейронном блоке нейропроцессора в импульсном формате представления данных (например, чисел), характерном для искусственных спайковых нейронных сетей, и которые наиболее приближены к биологическим нейронным системам [12]. При работе выходного блока отсутствует необходимость в операциях сжатия и фильтрации информационного потока, которые выполняются во входном устройстве нейропроцессора. Таким образом, основной операцией выходного устройства является перекодировка формата данных спайковых нейронных сетей в стандартное цифровое представление.

### 5.6.2. Преобразование частоты импульсов от одного нейрона

Обработка импульсных сигналов возможна в логической матрице [2], поскольку один слой матрицы представляет собой набор логических вентилей И или ИЛИ с произвольно подключаемыми входами. Путем маршрутизации

импульсных сигналов, с объединением их по логике И–ИЛИ на одной линии возможно задавать информационную величину или ее модифицировать. На рис. 5.25 приведены примеры модификации частотного сигнала с помощью логических элементов. Логический элемент ИЛИ осуществляет увеличение частоты при объединении импульсов входного сигнала с импульсами другого сигнала или генератора. Это эквивалентно операциям суммирования информационных величин или константы.

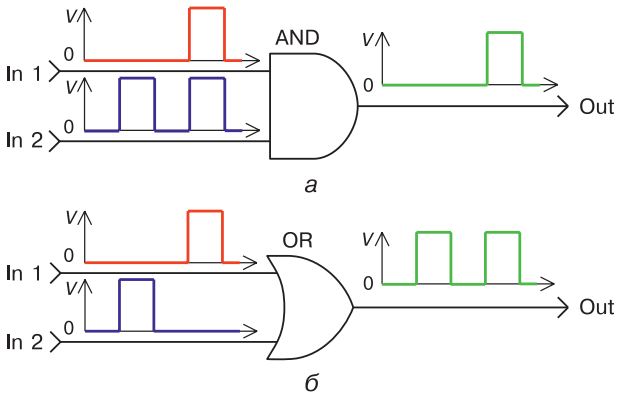



**Рис. 5.25.** Принцип модификации частотной информации в логических вентилях:

*а* — увеличение частоты информационных импульсов с помощью логического элемента ИЛИ (OR); *б* — уменьшение частоты информационных импульсов с помощью логического элемента И (AND)

Использование логической операции И позволяет уменьшать частоту вычитанием несовпадающих импульсов по времени, пропуская дальше модифицированный информационный сигнал. Добавление инверсии к логике И будет соответствовать операции вычитания информационных величин или вычитания константы.

### 5.6.3. Маршрутизация импульсов от популяции нейронов

На рис. 5.26 показан принцип модификации данных, представленных в популяционном коде. Модификация информации выполняется коммутируемой маршрутизацией информационных импульсов между позициями линий с помощью логической матрицы [2]. В этом случае возможна любая формула преобразования импульсов между входом INPUT и выходом OUTPUT.

Формула преобразования задается программируемыми связями матрицы маршрутизации. Внутри матрицы маршрутизации должно быть $N \cdot K$ возможных связей между $N$ входными и $K$ выходными линиями. Из них потребуется запрограммировать $K$ связей маршрутизации. Для взаимно

однозначного преобразования должно выполняться условие $N = K$ в случае биекционного отображения входного множества значений на выходное, и может быть $N > K$ в случае неполного преобразования.

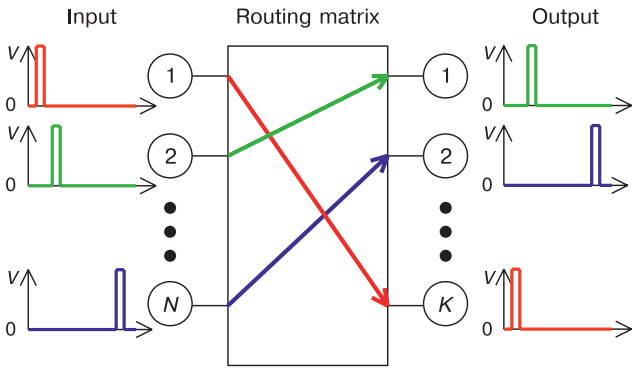



**Рис. 5.26.** Модификация информации путем маршрутизации импульсов от популяции нейронов

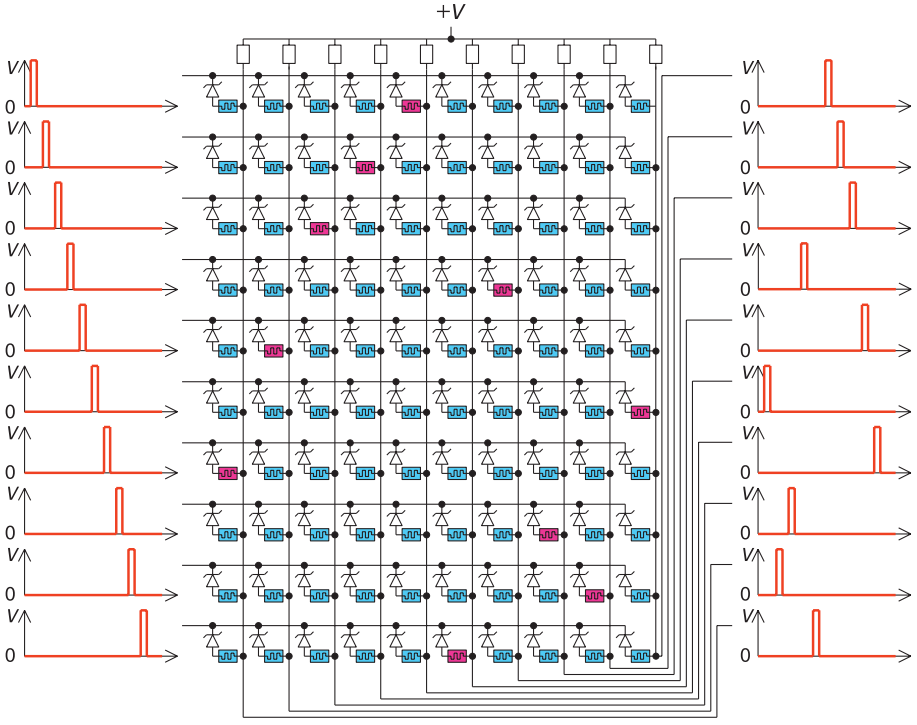



**Рис. 5.27.** Модификация информации путем маршрутизации импульсов в слое логической матрицы на основе комбинированного мемристорно-диодного кроссбара

В описанном принципе обработки информационных данных основное внимание заслуживает его энергоэффективность и быстродействие. Пространственное разделение информационных импульсов позволяет за два переключения логического элемента выполнять модификации над данными с программируемыми коэффициентами. Маршрутизация популяционного кода наиболее удобна для проведения математических операций с программируемыми константными величинами. Работу матрицы маршрутизации можно назвать распределенной, поскольку она будет заключаться всего лишь в нескольких импульсах на электрической связи в группе шин одной обрабатываемой величины.

На рис. 5.27 представлена принципиальная схема фрагмента логической матрицы на основе комбинированного мемристорно-диодного кроссбара, выполняющего маршрутизацию импульсов [38]. Красным цветом помечены мемристоры, запрограммированные в высокопроводящее состояние. Показанные на схеме диаграммы демонстрируют выполнение маршрутизации позиций входных импульсов, периодически поступающих на вход матрицы слева. На выходе справа на диаграммах видны перестановки позиций импульсов, что можно интерпретировать как операцию биекционного отображения последовательно поступающих входных данных.

### 5.6.4. Пространственно-временное преобразование информации

Преобразование выходных импульсов от одного нейрона является модификацией во времени, а маршрутизация импульсов от популяции нейронов — модификацией в пространстве. Объединение временного и пространственного способов модификации позволяет найти баланс между скоростью обработки и количеством задействованных элементов логической матрицы. Так можно производить операции над величинами, представленными в многоразрядном виде. В этом случае значения каждого разряда будут разделены пространственно, а сами разряды — во времени.

На рис. 5.28 показана блок-схема, отражающая идею поразрядной операции над десятичным трехразрядным числом, которая реализована на трех перестановочных матрицах маршрутизации $RM_1$–$RM_3$ (см. рис. 5.26). Время обработки составляет 3 такта. В каждом такте проводятся биекционные модификации с последовательно поступающими разрядами числа, закодированного в последовательность импульсов. Для каждого такта операции используется своя матрица маршрутизации. Матрицы маршрутизации установлены параллельно, и тактируемое устройство подключает в свой такт соответствующую матрицу маршрутизации, отключая при этом другие. В этом случае входной импульсный сигнал разделяется в пространстве и во времени. Импульсы на выходе объединяются по логике ИЛИ также в трехразрядное число в пространственно-временном формате представления.

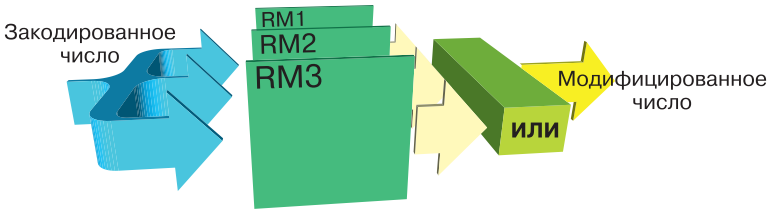



**Рис. 5.28.** Принцип модификации числа, закодированного в пространственно-временную последовательность импульсов

Пример реализации пространственно-временного преобразования на базе логической матрицы [2] показан на рис. 5.29.

Выходные данные представляют собой двухразрядное восьмеричное число. Сигналы на рис. 5.29 получены в результате SPICE-моделирования. На линию in3 для примера в качестве входной информации в пространственно-временном формате подаются два импульса, что соответствует восьмеричному числу $33_8$ (соответствует $27_{10}$). С помощью двух импульсов на линиях $InSD_1$ и $InSD_2$ выделяются поочередно разряды входного числа, которые подаются на маршрутизирующие матрицы, расположенные в первом слое 3D мемристорной логической матрицы. Входные импульсы подаются на следующий слой 3D логической матрицы без изменения позиции значений, но разделенные по разрядам. Во втором слое запрограммированный мемристорный слой выполняет двоичное преобразование позиций импульсов по логике ИЛИ с перестановочной функцией. После симуляции импульсы выходного числа остаются, разделенными по времени и разрядам нового значения, равного $35_8$ (соответствует $29_{10}$).

Таким образом, пространственно-временное представление информации позволяет производить последовательные поразрядные операции над числами. Это значительно расширяет диапазон дискретных значений обрабатываемых величин при незначительно больших временных затратах и меньших аппаратных затратах по сравнению с только пространственным представлением информации. В качестве недостатка использования в нейропроцессоре пространственно-временного представления информации можно отметить низкую помехозащищенность обрабатываемых данных. Потеря даже одного импульса на линиях мемристорных матриц приводит к сильному искажению обрабатываемой информации. Рассмотренное схемотехническое решение наилучшим образом подходит для применения в выходном блоке нейропроцессора для реализации преобразования формата представления выходной информации нейропроцессора в сжатом виде. Информация в сжатом виде выводится из схемы нейропроцессора с помощью минимального количества выходных электрических линий.

Стоит отметить, что рассмотренное пространственно-временное представление информации в выходном устройстве отличается отсутствием избыточности от кодирования информации популяцией нейронов во входном устройстве.

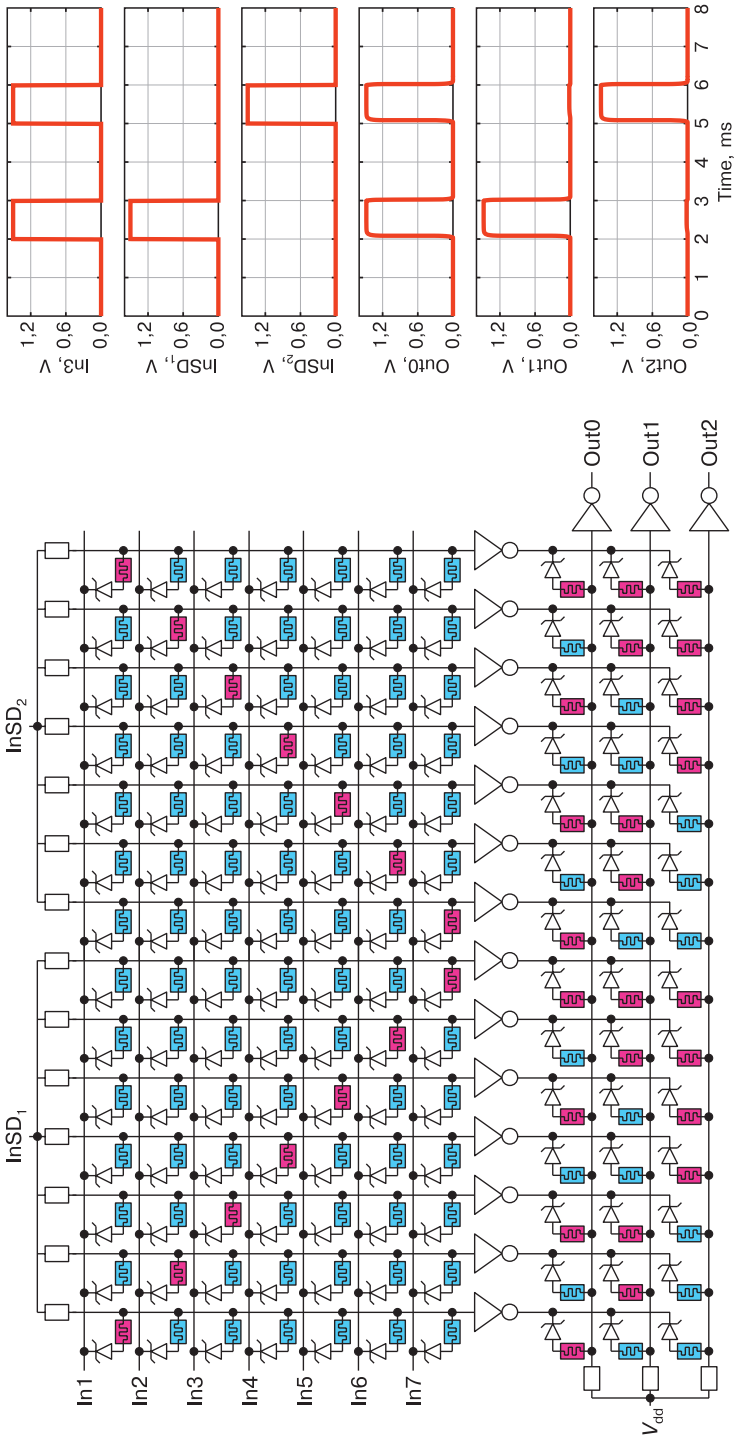



**Рис. 5.29.** Принципиальная схема модификации двухразрядного числа с помощью двух слоев логической матрицы и результаты ее SPICE-моделирования

### 5.6.5. Результаты SPICE-моделирования схем, декодирующих импульсные сигналы от популяции нейронов

В задаче декодирования импульсов от популяции нейронов в стандартные двоичные сигналы, решаемой выходным устройством нейропроцессора, можно предложить два решения: с использованием генераторов единичных импульсов и двоичных чисел соответственно. Оба решения построены с применением логической матрицы на основе комбинированного мемристорно-диодного кроссбара [38].

На рис. 5.30 показана принципиальная схема для преобразования популяционно кодированного числа в восьмеричное число, каждый разряд которого представлен бинарным кодом. Электрическая схема преобразователя реализуется в двух слоях логической матрицы с применением генератора единичных импульсов. Величины задержки двух входных импульсов, приходящих из нейронной сети, декодируются в двухразрядное восьмеричное число, значения разрядов которого представлены трехбитными двоичными числами. Входные импульсы по линиям Input 0 и Input 1 подаются на шины первого слоя матрицы. В этом слое выполняется поразрядное преобразование временных задержек импульсов в позиционный код. Во втором нижнем слое производится преобразование из позиционного кода в бинарный код. С помощью регистров и логических элементов «И», которые установлены на выходной периферии 3D логической матрицы, производится временная коррекция выходных импульсов.

Декодированное значение можно определить по диаграмме SPICE-моделирования, показанной на рис. 5.27 справа. На диаграммах для Input 0 и Input 1 импульсы показаны красным цветом. Определение задержки входных сигналов выполняется относительно импульсов от генераторов. Сигнал, показанный синей кривой, получен объединением импульсов внешних генераторов 1–7 на диаграмме. В первом фрейме положение входного импульса на линии Input 0 соответствует значению задержки 6. Это означает, что передаваемая величина имеет значение 6 в младшем разряде. Счет импульсов производится с конца фрейма. Также для импульса на линии Input 1 определяется задержка, равная 4, соответствующая старшему разряду передаваемой величины. Таким образом, в первом фрейме закодировано восьмеричное число $46_8$. Аналогично можно определить значение, передаваемое величины во втором фрейме. Оно имеет значение $33_8$. Восьмеричный формат представления выбран для упрощения принципиальной схемы и дальнейшего преобразования. Таким же образом можно передавать величины в других форматах, имеющих большее количество значений в разрядах. В результате преобразования на выходных линия lr_0—lr 2 и hr 0–hr 2 можно видеть импульсные сигналы в стандартном формате, соответствующие значениям $46_8$ и $33_8$.

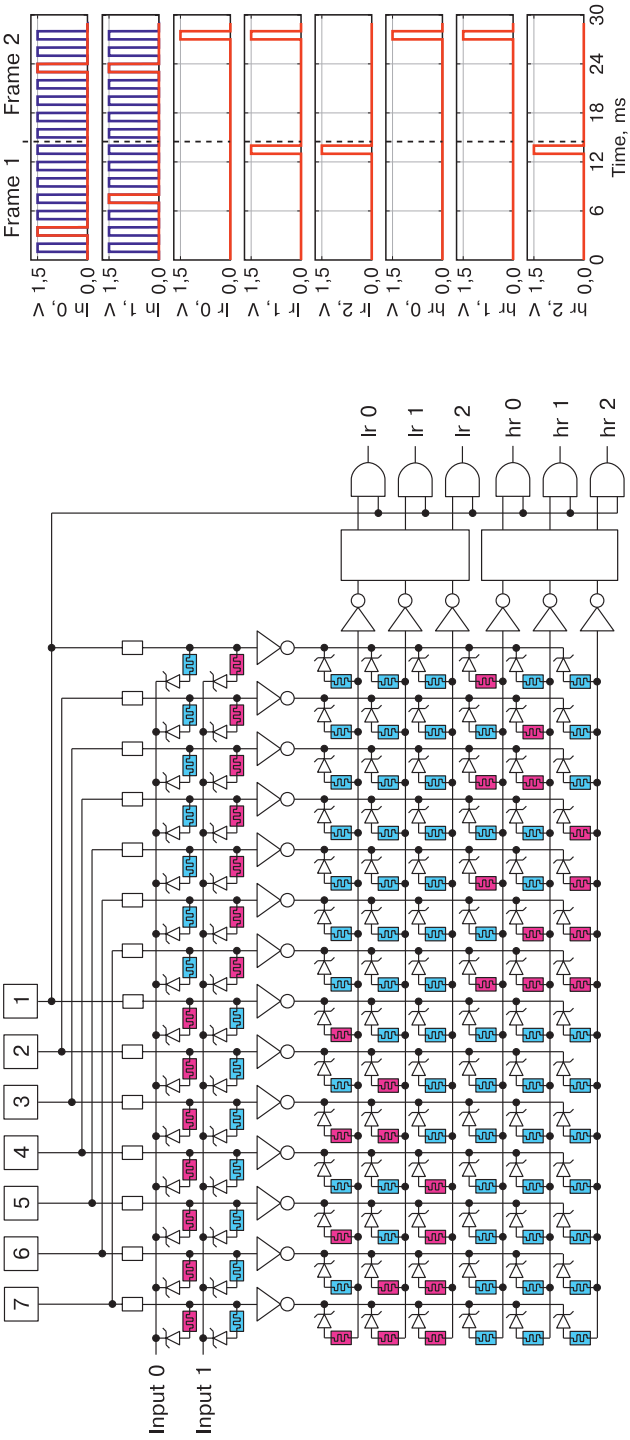



**Рис. 5.30.** Принципиальная электрическая схема декодирования популяционного двухразрядного импульсного сигнала в восьмеричный двухразрядный код и результаты ее SPICE-моделирования

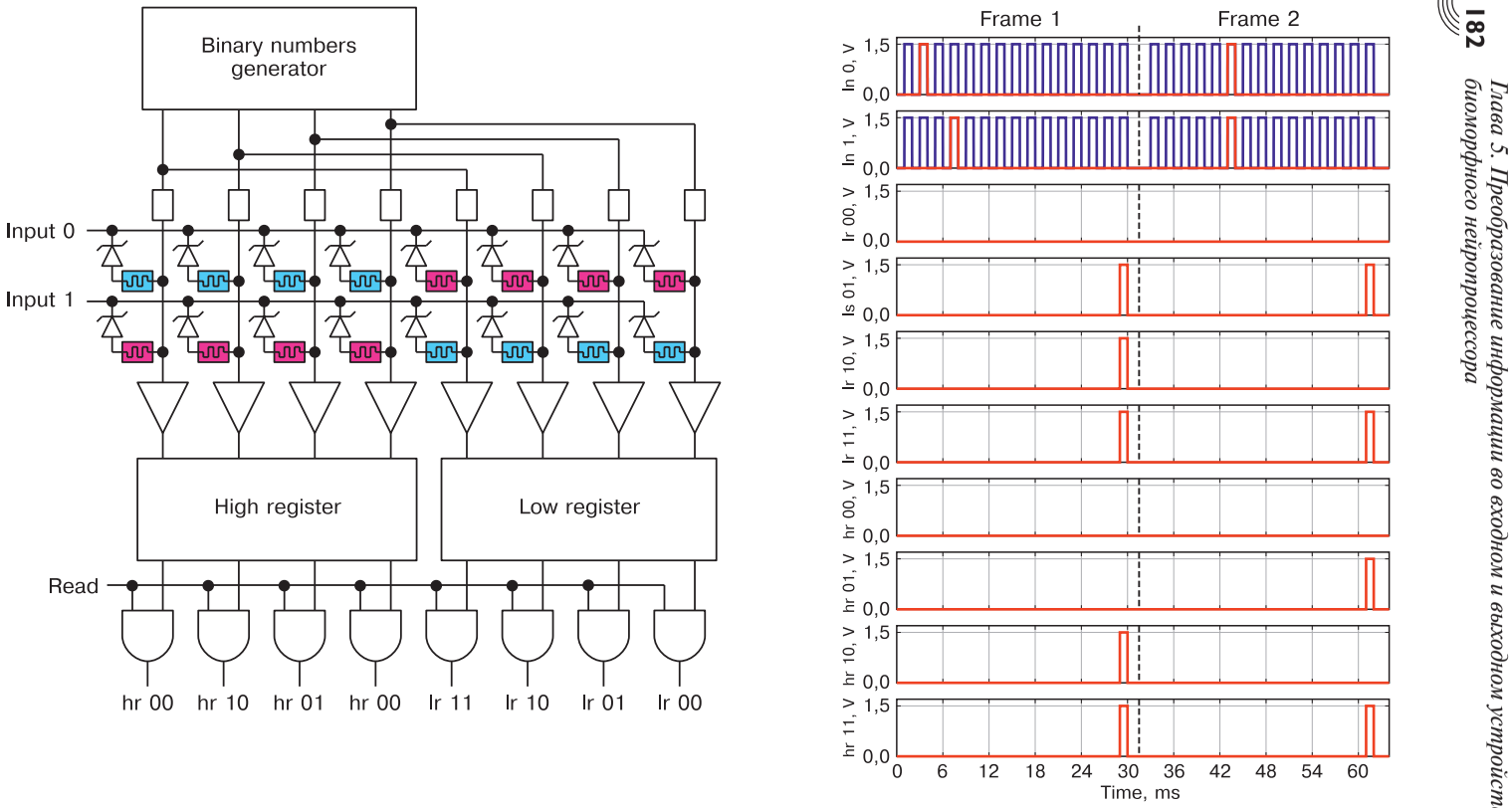



**Рис. 5.31.** Принципиальная электрическая схема декодирования популяционного двухразрядного импульсного сигнала в шестнадцатеричный двухразрядный код с использованием генератора бинарных чисел и результаты ее SPICE-моделирования

Другое, компактное схемотехническое решение той же задачи, но с использованием генератора двоичных чисел показано на рис. 5.31. Компактность схемы достигается за счет того, что наличие генератора двоичных чисел позволяет избавиться от преобразования позиционного кода в двоичный. Схемотехнически это означает отсутствие нижней логической матрицы на рис. 5.30.

Для примера в качестве выходного стандартного сигнала выбрано шестнадцатеричное двухразрядное число, представленное стандартным байтом информации в параллельном коде. Значения преобразуемой величины на входе показаны на диаграммах SPICE-моделирования сигналами Input 0 и Input 1. Определение закодированной величины производится аналогично примеру, показанному на рис. 5.30, с тем отличием, что максимальное значение разряда увеличено до 15 для шестнадцатеричного представления.

Значения преобразуемой величины на входе показаны на диаграммах SPICE-моделирования сигналами Input 0 и Input 1. Определение закодированной величины производится аналогично примеру, показанному на рис. 5.27, с тем отличием, что максимальное значение разряда увеличено до 15 для шестнадцатеричного представления.

Оригинальность работы устройства заключается в коммутации логической матрицей сигналов генератора на выход нейропроцессора на основе временной задержки входного импульса из аппаратной нейронной сети. Использование мемристорной логической матрицы во всех узлах нейропроцессора, включая выходное устройство, позволяет унифицировать элементную базу полной электрической схемы нейропроцессора, а также источников ее электропитания.

Список литературы

1. *Udovichenko S.Yu.*, *Pisarev A.D.*, *Busygin A.N.*, *Maevsky O.V.* Neuroprocessor based on combined memristor-diode crossbar // Nanoindustry. 2018. No. 5. Pp. 344–355.
2. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor — diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
3. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Yu.*, *Maevsky O.V.* 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor // Microelectronic Engineering. 2018. Vol. 198. Pp. 1–7.
4. *Gonzalez R.*, *Woods R.* The world of digital processing. Digital image processing // Transl. from English; P.A. Chochia (ed.). Moscow: Technosphere, 2005. P. 1072.
5. *Писарев А.Д.* Реализация дискретного косинусного преобразования во входном блоке мемристорного нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. № 1. С. 147–161.
6. *Li C.*, *Hu M.*, *Li Y.* et al. Analogue signal and image processing with large memristor crossbars // Nature electronics. 2018. Vol. 1. No. 1. Pp. 52–59.
7. *Levy Y.*, *Bruck J.*, *Cassuto Y.* et al. Logic operations in memory using a memristive Akers array // Microelectronics Journal. 2014. Vol. 45. Pp. 1429–1437.

8. *Hodgkin A.L.*, *Huxley A.F.* A quantitative description of membrane current and its application to conduction and excitation in nerve // Journal of Physiology. 1952. Vol. 117. No. 4. Pp. 500–544.
9. *Takeuchi T.*, *Duszkiewicz A.J.*, *Morris R.G.M.* The synaptic plasticity and memory hypothesis: Encoding, storage and persistence // Philosophical Transactions of the Royal Society B Biological Sciences. 2014. Vol. 369. No. 1633. P. 20130288.
10. *Gardner B.*, *Grüning A.* Supervised learning in spiking neural networks for precise temporal encoding // PLoS ONE. 2016. Vol. 11. No. 8. Pp. 1–28.
11. *Johansson R.S.*, *Birznieks I.* First spikes in ensembles of human tactile afferents code complex spatial fingertip events // Nature Neuroscience. 2004. Vol. 7. No. 2. Pp. 170–177.
12. *Gollisch T.*, *Meister M.* Rapid neural coding in the retina with relative spike latencies // Science. 2008. Vol. 319. No. 5866. Pp. 1108–1111.
13. *Mainen Z.F.*, *Sejnowski T.J.* Reliability of spike timing in neocortical neurons // Science. 1995. Vol. 268. No. 5216. Pp. 1503–1506.
14. *Reich D.S.*, *Victor J.D.*, *Knight B.W.* et al. Response variability and timing precision of neuronal spike trains in vivo // Journal of Neurophysiology. 1997. Vol. 77. No. 5. Pp. 2836–2841.
15. *Uzzell V.*, *Chichilnisky E.* Precision of spike trains in primate retinal ganglion cells // Journal of Neurophysiology. 2004. Vol. 92. No. 2. Pp. 780–789. pmid:15277596.
16. *Bloom F.*, *Leiserson A.*, *Hofstedter L.* Brain, mind and behavior / Transl. from English. Moscow: Mir, 1988. 248 p.
17. *Van Rullen R.*, *Guyonneau R.*, *Thorpe S.J.* Spike times make sense // Trends in Neurosciences. 2005. Vol. 28. No. 1. Pp. 1–4.
18. *Larkum M.E.*, *Zhu J.J.*, *Sakmann B.* Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons // Journal of Physiology. 2001. Vol. 533. No. 2. Pp. 447–466.
19. *Gütig R.* To spike, or when to spike? // Current Opinion in Neurobiology. 2014. Vol. 25. Pp. 134–139.
20. *Kasinski A.*, *Ponulak F.* Comparison of supervised learning methods for spike time coding in spiking neural networks // International Journal of Applied Mathematics and Computer Science. 2006. Vol. 16. No. 1. Pp. 101–113.
21. *Mohemmed A.*, *Schliebs S.*, *Matsuda S.*, *Kasabov N.* SPAN: Spike pattern association neuron for learning spatio-temporal spike patterns // International Journal of Neural Systems. 2012. Vol. 22. No. 4. P. 1250012.
22. *Yu Q.*, *Tang H.*, *Tan K.C.*, *Li H.* Precise-spike-driven synaptic plasticity: Learning hetero-association of spatiotemporal spike patterns // PLoS ONE. 2013. Vol. 8. No. 11. P. e78318. pmid:24223789.
23. *Писарев А.Д.* Энергоэффективное импульсное кодирование входной информации для нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. № 3. С. 186–212.
24. *Писарев А.Д.*, *Маевский О.В.*, *Бусыгин А.Н.*, *Удовиченко С.Ю.* Многослойная логическая матрица на основе мемристорной коммутационной ячейки. 2019. Патент № 2682548.
25. *Biolek D.*, *Di Ventra M.*, *Pershin Y.V.* Reliable SPICE Simulations of Memristors, Memcapacitors and Meminductors // Radioengineering. 2013. Vol. 22. No. 4. Pp. 945–968.
26. *Писарев А.Д.*, *Бусыгин А.Н.*, *Бобылев А.Н.*, *Удовиченко С.Ю.* Комбинированный мемристорно-диодный кроссбар как основа запоминающего устройства // Вестник ТюмГУ.Физико-математическое моделирование. Нефть, газ, энергетика. 2017. Т. 3. № 4. С. 142–149.

27. *Pisarev A.D.*, *Busygin A.N.*, *Bobylev A.N.*, *Udovichenko S.Yu.* Operation principle and fabrication technology of the neuroprocessor input unit on the basis of the memristive logic matrix // International Journal of Nanotechnology. 2019. Vol. 16. No. 6–10. Pp. 596–601.
28. *Писарев А.Д.* Spice-моделирование процессов ассоциативного самообучения и безусловного разобучения в логическом блоке нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2018. № 3. С. 132–145.
29. *Lobo J.L.*, *Ser J.D.*, *Bifet A.*, *Kasabov N.* Spiking Neural Networks and online learning: An overview and perspectives // Neural Networks. 2020. Vol. 121. Pp. 88–100.
30. *Thorpe S.J.*, *Guyonneau R.*, *Guilbaud N.* et al. Spike net: Real-time visual processing with one spike per neuron // Neurocomputing. 2004. Vol. 58–60. Pp. 857–864.
31. *Pan Z.*, *Wu J.*, *Zhang M.*, *Li H.*, *Chua Y.* Neural population coding for effective temporal classification // 2019 International Joint Conference on Neural Networks (IJCNN). 2019. arXiv:1909.08018v2
32. *Ponulak F.*, *Kasinski A.* Introduction to spiking neural networks: Information processing, learning and applications // Acta Neurobiologiae Experimentalis. 2011. Vol. 71. No. 4. Pp. 409–433.
33. *Lobov S.*, *Mironov V.*, *Kastalskiy I.*, *Kazantsev V.* A spiking neural network in sEMG feature extraction // Sensors. 2015. Vol. 15. No. 11. Pp. 27894–27904.
34. *Chan V.H.*, *Carey R.M.* Simultaneous latency and rate coding for automatic error correction. 2019. US Patent 10282660.
35. *Thakur C.S.*, *Hamilton T.J.*, *Wang R.* et al. A neuromorphic hardware framework based on population coding // 2015 International Joint Conference on Neural Networks (IJCNN). 2015. arXiv:1503.00505.
36. *Nuno-Maganda M.*, *Torres-Huitzil C.* A temporal coding hardware implementation for spiking neural networks // ACM SIGARCH Computer Architecture News. 2011. Vol. 38. No. 4. P. 2.
37. *Eurich C.W.*, *Wilke S.D.* Multidimensional encoding strategy of spiking neurons // Neural Computation. 2000. Vol. 12. No. 7. Pp. 1519–1529.
38. *Писарев А.Д.*, *Бусыгин А.Н.*, *Ибрагим А.Х.*, *Удовиченко С.Ю.* Преобразование информации в выходном устройстве биоморфного нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2020. Т. 6. № 4 (Принята в печать. Выпуск в декабре 2020).

# ГЛАВА 6

# СОЗДАНИЕ АППАРАТНОЙ ОСНОВЫ БИОМОРФНОГО НЕЙРОПРОЦЕССОРА

## 6.1. ОБОРУДОВАНИЕ ДЛЯ ИЗГОТОВЛЕНИЯ И ИССЛЕДОВАНИЯ НАНОМАТЕРИАЛОВ И ЭЛЕКТРОННЫХ УСТРОЙСТВ НА ИХ ОСНОВЕ

Важной тенденцией развития микро- и наноэлектроники в современных условиях является интеграция технологических процессов, которая реализуется с помощью технологических кластерных установок. Разработанная группой компаний НТ-МДТ платформа «НаноФаб-100» представляет собой несколько технологических модулей, окружающих центрального транспортного робота. Транспортный модуль передает обрабатываемые образцы последовательно из одного модуля в другой в условиях вакуума.

Организация технологических модулей в кластеры по требованиям к чистоте и к вакууму позволяет сократить до минимума временные потери на загрузку/выгрузку образцов и регламентные работы, а сквозная транспортная система с точностью позиционирования до 5 микрон обеспечивает конвейерный характер всех производимых технологических операций. Кроме этого, кластерная организация технологических модулей дает возможность пристыковки всей системы к другим технологическим и исследовательским комплексам, например к станции синхротронного излучения.

На современном этапе при производстве интегральных схем на кристалле перед технологами и разработчиками технологического оборудования стоит задача создания полностью замкнутой и интегрированной в пространстве системы — суперкластера (микрофабрикатора). Основными препятствиями на пути решения данной проблемы являются жидкостные процессы нанесения и проявления фоторезистивных масок, химико-механической планаризации функциональных слоев и химической очистки поверхности подложек. Поэтому ведущие зарубежные фирмы проводят интенсивные исследования с целью замены этих процессов на «сухие», проводимые в вакууме с использованием лазерных и ионных пучков и газоразрядной плазмы. В этом случае удастся решить проблему реализации полностью «сухого» технологического процесса изготовления ИС на основе

суперкластера. При этом вопросы о количестве и номенклатуре модулей должны решаться в процессе моделирования работы суперкластера в различных режимах.

На рис. 6.1 представлен кластерный нанотехнологический комплекс (НТК) «НаноФаб-100» компании NT–MDT, установленный в Научно-образовательном центре (НОЦ) «Нанотехнологии» Тюменского государственного университета и предназначенный для создания и исследования наноматериалов и устройств для микро- и наноэлектроники. Такой нанотехнологический комплекс не дает возможности создавать интегральную схему (ИС) за один вакуумный цикл. Маски для напыления / травления микросхемы приходится изготавливать отдельно с помощью электронной литографии на электронном микроскопе JSM-6510LV-EDS.


**Рис. 6.1.** Внешний вид НТК «НаноФаб-100»

НТК «НаноФаб-100» включает транспортный модуль и четыре технологических модуля: плазмохимического травления, магнетронного распыления, травления ионным пучком и ионной имплантации. Обслуживание этих модулей, программирование и контроль технологических процессов происходят в автоматическом режиме с помощью персонального компьютера.

Очистка поверхности подложки проводится в модуле плазмохимического травления (ПХТ) (рис. 6.2), который предназначен для «сухого травления» и очистки поверхности металлов и диэлектриков. Сочетание двух источников плазмы (емкостного и индуктивного) позволяет получить плазму с плотностью химически активных ионов фтора $F^+$, необходимой для создания структур с высоким аспектным соотношением, и реализовывать режим как анизотропного, так и изотропного травления.


**Рис. 6.2.** Модуль плазмохимического травления (ПХТ)

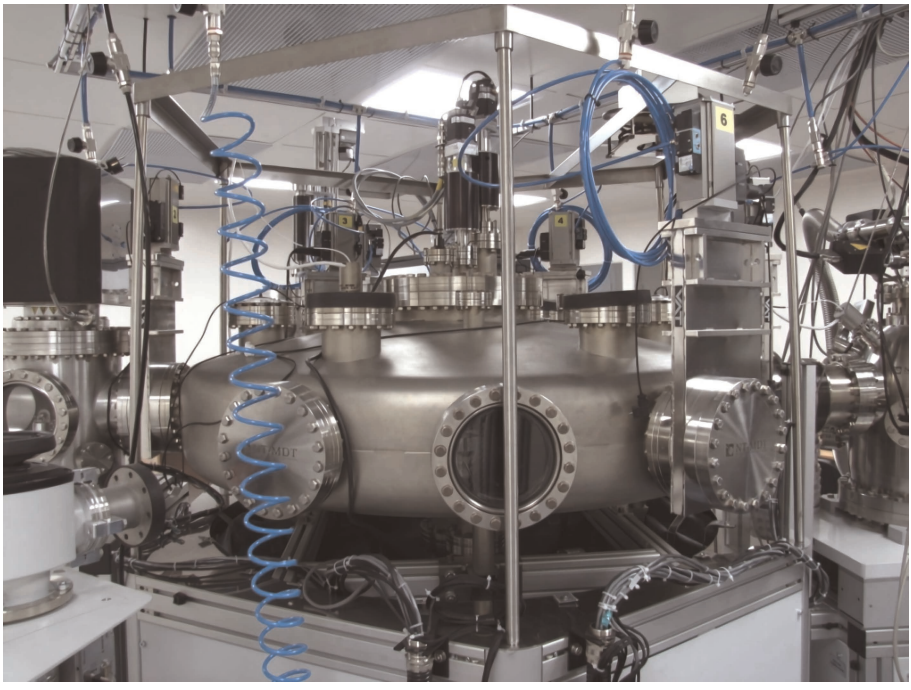



**Рис. 6.3.** Радиальный транспортный модуль

Очищенная подложка с помощью сверхвысоковакуумного радиального транспортного модуля (рис. 6.3) перемещается в магнетронный модуль.

Основой транспортного модуля является робот-раздатчик, предназначенный для передачи образцов между модулями. Рабочее давление составляет $4 \times 10^{-10}$ тор.

В магнетронном модуле (рис. 6.4) распыление металлической мишени (катода) при постоянном токе и диэлектрической мишени при импульсном токе обеспечивает нанесение на подложку однослойных и многослойных наноструктурированных пленок с заданным составом.

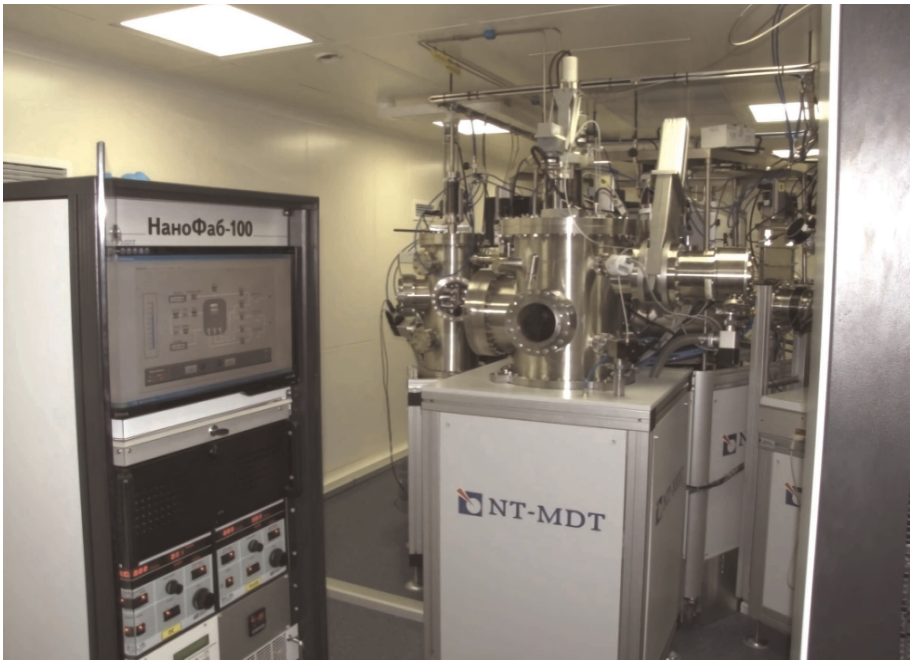



**Рис. 6.4.** Модуль магнетронного нанесения пленок и покрытий

Далее подложка с нанесенной пленкой с помощью радиального транспортного модуля передается в сверхвысоковакуумный модуль с ускоряющей колонной фокусированных ионных пучков с энергией до 30 КэВ (рис. 6.5), предназначенный для проведения технологических операций ионного травления поверхности материалов, нанолитографии, резки и визуализации наноэлементов и наноструктур, очистки поверхности полупроводниковых пластин и пр.

С помощью находящейся в модуле системы вторично-ионной масс-спектроскопии (ВИМС) при послойном ионном распылении (травлении) проводился анализ распределения элементов по толщине тонкой пленки смешанного оксида металлов $Ti_{0,85}Al_{0,15}O_{1,93}$, предназначенной для изготовления мемристорного кроссбара [1].

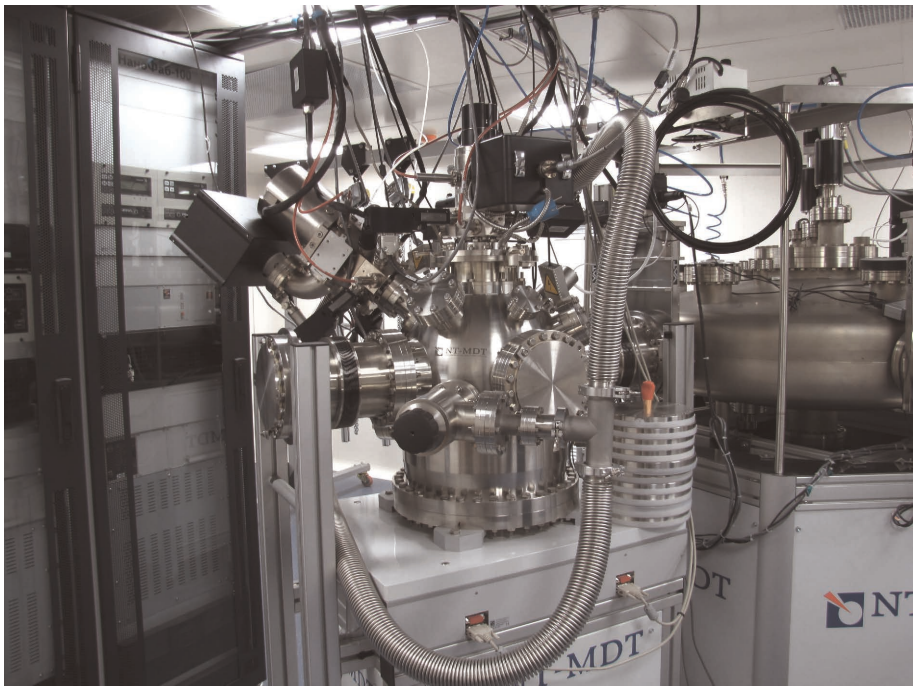



**Рис. 6.5.** Сверхвысоковакуумный модуль нанообработки фокусированными ионными пучками

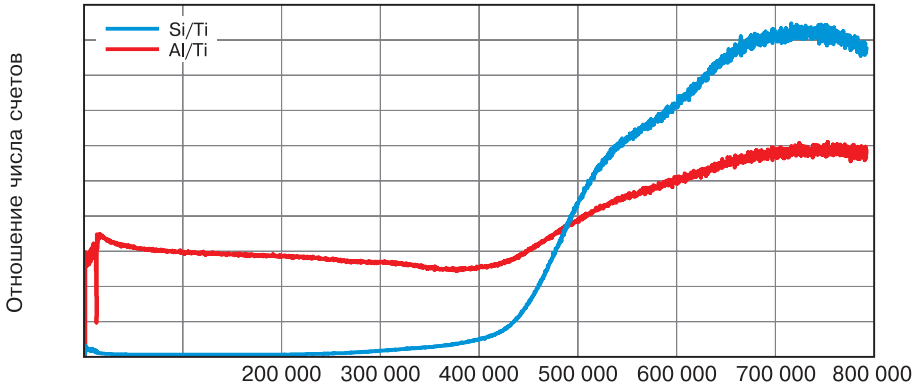



**Рис. 6.6.** Изменение отношения концентраций ионов алюминия и титана в процессе распыления ионным пучком пленки с 15 % содержания алюминия

На рис. 6.6 показано изменение отношения концентрации алюминия и титана, а также кремния и титана в процессе распыления пленки ионным пучком. Резкое увеличение концентрации кремния в правой части графика

свидетельствует об окончании процесса распыления пленки и начале распыления слоя оксида кремния на подложке. Отношение концентраций ионов титана и алюминия по толщине пленки изменялось от 7,44 до 6,36.

Проводящие дорожки шириной 200–400 нм наносились в магнетронном модуле на подложку и мемристорную пленку через маски, приготовленные на сканирующем электронном микроскопе JSM-6510LV-EDS (рис. 6.7).

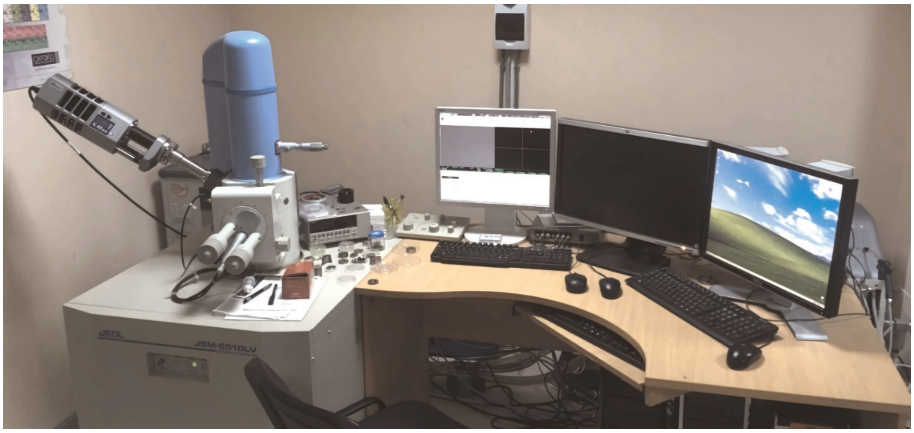



**Рис. 6.7.** Электронный микроскоп JSM-6510LV с энергодисперсионной приставкой Oxford Instruments

Литографическая приставка NanoMaker Full к этому электронному микроскопу позволяет изготавливать маски для изготовления мемристорных кроссбаров и для плазмохимического травления интегральных микросхем с разрешением не хуже 20 нм.

С точки зрения единого технологического цикла для изготовления полупроводниковых пленок диода Зенера, так же как и для мемристорной пленки, применялся магнетронный модуль с мишенями (катодами) из кремния, легированного бором и фосфором. Вспомогательным является модуль ионной имплантации (рис. 6.8), который позволяет сравнивать электрофизические свойства диодов, изготовленных методами магнетронного распыления и ионной имплантации.

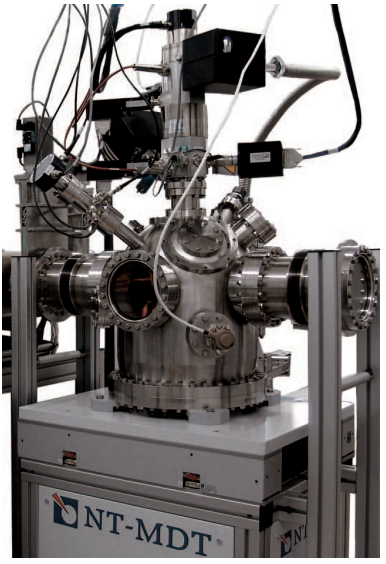



**Рис. 6.8.** Имплантационный модуль фокусированных ионных пучков с энергией до 30 КэВ

## 6.2. ИЗГОТОВЛЕНИЕ МЕМРИСТОРОВ С ВЫСОКИМИ ЭЛЕКТРИЧЕСКИМИ ХАРАКТЕРИСТИКАМИ НА ОСНОВЕ СМЕШАННЫХ ОКСИДОВ МЕТАЛЛОВ

### 6.2.1. Выбор мемристорного материала

В мемристоре между предельными высокопроводящим и низкопроводящим состояниями имеется множество промежуточных состояний с разной проводимостью. Эти состояния можно использовать в процессах ассоциативного самообучения нейросети на основе мемристорных синапсов и одновременной обработки входных импульсов, заключающейся в их взвешивании и суммировании в нейропроцессоре [2].

Чем шире диапазон резистивного переключения мемристора, тем больше можно реализовать синаптических связей с помощью этого мемристора в нейропроцессоре. Например, в мемристоре на основе пленки диоксида титана $TiO_2$ толщиной 30 нм и платиновыми электродами [3] отношение сопротивлений в закрытом $R_{off}$ и открытом $R_{on}$ состояниях при напряжении считывания 0,2 В имеет величину $R_{off}/R_{on} \approx 2$. Чистый оксид циркония в структуре $Ti/ZrO_2/Pt$ дает $R = 10^4$, а гафния в $Pt/HfO_2/HfO_{2-x}/TiN$, соответственно, $10^3$ [4].

Однако такие мемристоры обладают нестабильностью напряжений переключения и предельных сопротивлений. Так, в работе [5] структура $Pt/HfO_2/TiN$ площадью $1{,}6 \cdot 10^3$ нм$^2$ и толщиной активного слоя 3 нм имеет максимальное отношение $R = 15$ при максимальных отклонениях $\Delta R_{on} = 85$ % и $\Delta R_{off} = 88$ %; в работе [6] структура Ti/ZnO/TiN с активным слоем толщиной 23 нм и площадью $10{,}4 \cdot 10^3$ нм$^2$ $\Delta R_{on} = 30$ %, $\Delta R_{off} = 17$ %; в работе [7] структура $Pt/ZrO_2/TiN$ с толщиной активного слоя 20 нм $\Delta R_{on} = 50{,}4$ %, $\Delta R_{off} = 68{,}5$ %. Такая нестабильность сопротивлений в указанных оксидах не дает возможности использовать их для построения больших мемристорных кроссбаров.

Проблему высокого разброса значений сопротивлений можно решить путем легирования оксида переходного металла. Добавка Al в среднем 50 % от концентрации Hf по толщине пленки $HfO_2$ [8] приводит к уменьшению $\Delta R_{on}$ до 66 % и $\Delta R_{off}$ до 49 %. В работе [7] теоретически и экспериментально показано, что замещение 3 % атомов циркония в $ZrO_2$ атомами алюминия Al приводит к снижению $\Delta R_{on}$ с 50,4 до 26,5 % и $\Delta R_{off}$ с 68,5 до 18,3 %. Наибольшая стабильность была достигнута в структуре $TiN/Ti_{0,92}Al0_{0,08}O_y/TiN$ [1] с толщиной активного слоя 20 нм при $\Delta R_{on} = 3{,}3$ % и $\Delta R_{off} = 15{,}7$ %.

Кроме этого, в мемристорных материалах на основе смешанных оксидов переходных металлов можно добиться увеличения диапазона переключения сопротивлений мемристоров по сравнению с мемристорами

на чистых оксидах. При добавлении примеси атомов алюминия в оксиды четырехвалентных переходных металлов: $Ti_xAl_{1-x}O_y$ [9], $Hf_xAl_{1-x}O_y$ [10], энергия связи ионов кислорода и основного металла уменьшается, что приводит к уменьшению порогового напряжения переключения и активизации процесса обеднения слоя ионами кислорода за счет их миграции в электрическом поле. Повышенная мобильность ионов кислорода приводит к увеличению отношения максимального сопротивления к минимальному.

### 6.2.2. Нанотехнология изготовления мемристорного устройства на основе смешанного оксида металлов

При получении пленки смешанного оксида металлов с помощью метода атомно-слоевого осаждения в ней остаются примеси реагентов. Примеси и неоднородное распределение элементов по толщине пленки увеличивают неоднородность электрического поля, что приводит к нестабильности электрических характеристик устройства [1].

В отличие от метода атомно-слоевого осаждения [9; 10] метод реактивного магнетронного осаждения смешанного оксида металлов дает возможность исключить примеси и получить высокую равномерность распределения элементов по толщине пленки, что приводит к большей стабильности порогового напряжения переключения мемристора и его сопротивлений в низкопроводящем и высокопроводящем состояниях [1].

Напыление эталонной пленки диоксида титана $TiO_2$ толщиной 30 нм проводилось в магнетронном модуле на постоянном токе, входящем в нанотехнологический комплекс (НТК) NT-MDT «НаноФаб-100» [11; 12]. Электромагнитные клапаны на газовых магистралях установки позволяют регулировать расход газов с точностью до 0,1 ссм$^3$/мин. Диаметр распыляемой мишени — 78 мм. Вакуумная подготовка реакторной камеры осуществлялась при давлении $5 \cdot 10^{-5}$ Па. Напыление осуществлялось в импульсном режиме магнетрона при постоянной мощности 125 Вт и постоянном давлении 0,25 Па. Расход аргона для поддержания рабочего давления составлял 22 ссм$^3$/мин, расход кислорода составлял 8 ссм$^3$/мин.

Мемристоры изготовлены по технологии кроссбар (рис. 6.9) путем последовательного напыления слоев через маски электронного резиста (PMMA), выполненные на электронном микроскопе JSM-6510LV-EDS.

Нижний электрод состоит из 5 нм адгезивного подслоя Ti и 30 нм слоя W. Активный слой — 30 нм $TiO_2$, $Ti_{0,93}Al_{0,07}O_x$, $Ti_{0,9}Al_{0,1}O_x$ и $Ti_{0,85}Al_{0,15}O_x$, соответственно в четырех экспериментах. Верхний электрод — 95 нм TiN. Ширина проводящих дорожек в месте пересечения составляет 1 мкм.

На рис. 6.10 представлены фотографии чипов мемристорных кроссбаров, в которых пленка смешанного оксида имеет разную долю алюминия.

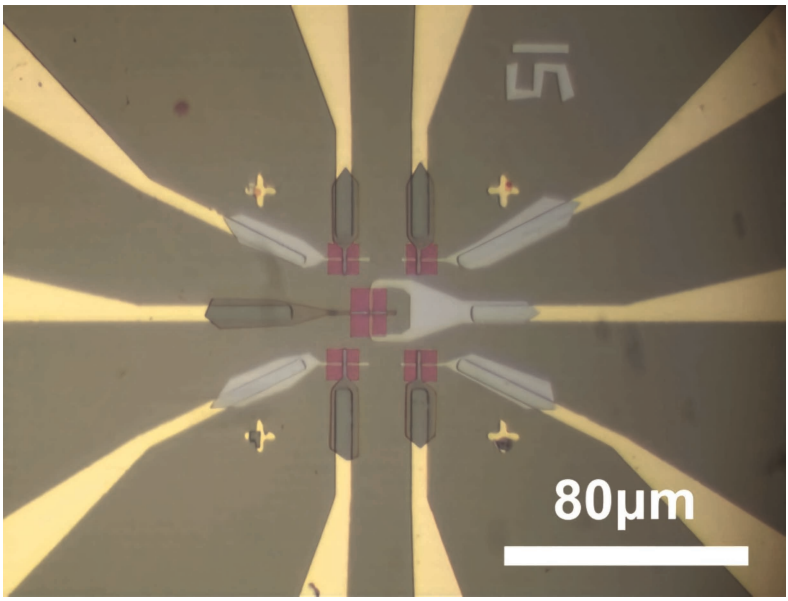



**Рис. 6.9.** Фотография с оптического микроскопа экспериментального образца мемристорного кроссбара, демонстрирующая совмещение микроразмерных и наноразмерных проводников

**Рис. 6.10.** Образцы мемристорных кроссбаров на основе пленки оксида титана. Цвет пленки зависит от доли примеси алюминия в оксиде

### 6.2.3. Метод получения смешанного оксида с контролируемым содержанием металлов

При распылении в магнетронном модуле одновременно с титаном второго катода из алюминия необходимо контролировать количественный состав выращиваемого слоя смешанного оксида металлов. Для этого были проведены следующие расчеты, результаты которых согласуются с экспериментальными данными.

Количественный состав пленки при постоянной скорости напыления должен оставаться однородным по толщине. Возле каждого катода в магнетроне имеется акустический датчик скорости напыления. Толщина напыленного на датчик слоя металла определяется выражением

$$d = \frac{m}{\rho S},$$

где $m$ — масса напыляемого вещества; $\rho$ — его плотность; $S$ — площадь датчика. Отношение скоростей напыления при одновременном распылении двух катодов с постоянными скоростями равно отношению толщины пленок, осаждаемых на датчиках с одинаковой площадью

$$\frac{d_1}{d_2} = \frac{m_1\rho_2}{m_2\rho_1}.$$

Молекула стехиометрического оксида алюминия $Al_2O_3$ содержит два атома алюминия, молекула оксида титана в максимальной степени окисления (полностью оксидный режим реактивного распыления) $TiO_2$ содержит один атом титана. Следовательно, отношения количества вещества оксидов и их металлов отличаются в два раза

$$\frac{v_{Ti}}{v_{Al}} = \frac{2v_{TiO_2}}{v_{Al_2O_3}} = \frac{2v_1}{v_2}.$$

Представляя массу как произведение количества вещества, молярной массы $M$ и постоянной Авогадро, получим для отношения толщины пленок на датчиках и, соответственно, для отношения скоростей распыления катодов следующее выражение:

$$D = \frac{d_1}{d_2} = 2\frac{\nu_1}{\nu_2}\frac{M_1}{M_2}\frac{\rho_2}{\rho_1}.$$

Отношения скоростей напыления составили $D$ (7 ат. % Al) = 19,4; $D$ (10 ат. % Al) = 13,1; $D$ (15 ат. % Al) = 8,28. Тестирование этой расчетной модели скорости распыления катодов проводилось при получении составов в образцах 1–3: $Ti_{0,9}Zr_{0,1}O_2$, $Ti_{0,8}Zr_{0,2}O_2$ и $Ti_{0,7}Zr_{0,3}O_2$ соответственно, для которых использовалось соотношение

$$\frac{v_{Ti}}{v_{Al}} = \frac{v_{TiO_2}}{v_{ZrO_2}} = \frac{v_1}{v_2}.$$

Полученные пленки исследовались на рентгеновском фотоэлектронном спектрометре (РФЭС) Thermo Fisher Scientific K-ALPHA (табл. 6.1).

Таким образом, предложенный метод определения скорости распыления катодов позволял достаточно точно контролировать количественный состав выращиваемого слоя смешанного оксида металлов.

*Таблица 6.1*

**Результаты измерений РФЭС**

| Образец | 1 | 2 | 3 |
|---|---|---|---|
| $v_{Ti}/v_{Zr}$ расчетное | 90/10 | 80/20 | 70/30 |
| $v_{Ti}/v_{Zr}$ измеренное | 84,07/15,93 | 79,89/20,11 | 75,51/24,49 |

Результаты исследования распределения мольной доли металлов по толщине пленки оксида титана с примесью циркония методом РФЭС представлены на рис. 6.11, *а*.

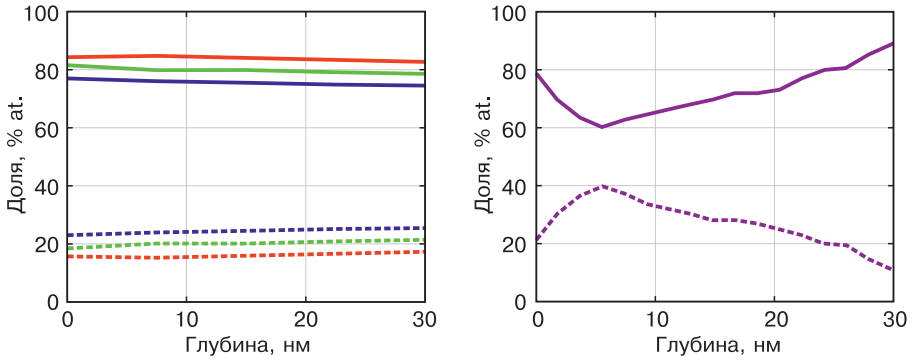



**Рис. 6.11.** Распределение мольных долей по толщине пленки:

*а* — Ti и Zr: красные кривые — образец 1, зеленый — образец 2, синий — образец 3 (сплошными кривыми показано содержание титана, а штрихованными кривыми — примеси); *б* — Ti и Al в работе [13] (сплошная кривая — Al)

Эти результаты свидетельствуют о том, что метод реактивного магнетронного осаждения смешанного оксида металлов путем одновременного распыления двух катодов дает возможность получить более равномерное распределение элементов по толщине пленки активного слоя по сравнению с методом атомно-слоевого осаждения, что требуется для повышения стабильности электрических характеристик мемристора.

В активном слое мемристоров — смешанном оксиде металлов одним из элементов может быть титан или цирконий, или гафний, а элементом примеси — трехвалентный металл с ионным радиусом, равным 0,7–1,2 ионного радиуса титана или циркония, или гафния соответственно [14]. Можно ожидать, что в мемристорах на смешанных оксидах $Ti_xSc_{1-x}O_y$, $Hf_xSc_{1-x}O_y$, $Hf_xY_{1-x}O_y$, $Hf_xLu_{1-x}O_y$, $Zr_xSc_{1-x}O_y$, $Zr_xY_{1-x}O_y$, $Zr_xLu_{1-x}O_y$, также будет наблюдаться оптимальная доля примеси, соответствующая максимально повышенному отношению сопротивлений в низкопроводящем и высокопроводящем состояниях.

## 6.3. ИЗГОТОВЛЕНИЕ КОМБИНИРОВАННОГО МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА — ОСНОВЫ АППАРАТНОЙ РЕАЛИЗАЦИИ НЕЙРОПРОЦЕССОРА

Результаты разработки электрической схемы, топологии и технологии изготовления комбинированного мемристорно-диодного кроссбара, необходимого для создания запоминающей и логической матрицы нейропроцессора, содержатся в [15; 16]. Выбор мемристорного материала и технология изготовления мемристорного устройства представлены в [11; 12].

Для изготовления лабораторного комбинированного мемристорно-диодного кроссбара остается подобрать материалы и технологию изготовления полупроводниковых слоев диода Зенера, обеспечивающие требуемые характеристики селективного элемента. Этими характеристиками являются: напряжение открытия в прямом смещении и при пробое, электрические сопротивления в закрытом и открытом состояниях, а также при обратимом пробое. Сопротивление диода в закрытом состоянии должно быть максимально возможным, а в открытом состоянии и при пробое — как можно меньшим. Высокое сопротивление в открытом состоянии приведет к падению напряжения на диоде, что в итоге потребует подачи большего напряжения для программирования мемристора в ячейке матрицы. Напряжение открытия диода при прямом смещении *p*-*n*-перехода должно быть минимальным, поскольку допустимая амплитуда информационных импульсов в логической матрице должна быть больше этой величины. Напряжение обратимого пробоя, соответственно, должно быть больше напряжения информационных импульсов, чтобы исключить изменение состояния мемристоров.

### 6.3.1. Выбор технологии изготовления полупроводниковых слоев диода Зенера

Существует несколько технологий изготовления полупроводникового диода, подходящего для использования в качестве селективного элемента комбинированного кроссбара.

Классический процесс термодиффузионного легирования кремния бором (для дырочной проводимости) и фосфором (для электронной проводимости) производится при температурах более 1000 °C. Легирование таким способом полупроводников селективного диода невозможно, поскольку при температуре выше 400 °C происходит разрушение нижележащих проводников, что показано в работах по 3D-интеграции германиевых транзисторов [17] и транзисторов на углеродных нанотрубках [18].

Электрохимическое осаждение (ECD) [19] пленки полупроводника *n*-типа — оксида цинка ZnO на легированную кремниевую подложку *p*-типа

с удельной проводимостью 0,03–0,05 Ом·см приводит к образованию структуры с выраженными выпрямляющими электрическими свойствами. Однако структура пленки получается существенно неоднородной в виде вертикальных столбиков со средним диаметром 200 нм и расстоянием между столбиками около 400 нм. Помимо этого, в обратной ветви вольт-амперной характеристики через *p*-*n*-переход на основе такой пленки не наблюдается обратимого электрического пробоя, необходимого для функционирования комбинированного кроссбара с биполярными мемристорами.

Гетеропереход, полученный спеканием под давлением полупроводниковых слоев $ZnO : Al_2O_3$ и $CuO : Li_2CO_3$ [20], обеспечивает обратимый электрический пробой при относительно высоком напряжении. При этом пленка оксида меди имеет поры диаметром в среднем 10 мкм, что неприемлемо для микроэлектронных устройств.

Метод ионного легирования [21] для изготовления кремниевого диода *p*-Si/*n*-Si более точный, надежный и воспроизводимый, чем метод термодиффузии. Однако дефекты после имплантации необходимо отжигать при 900–1100 °C [22], что для многослойных структур неприемлемо.

Получение морфологически более совершенных пленок обеспечивает метод реактивного магнетронного распыления. В работе [23] изготовлен кроссбар с униполярным мемристором на основе смешанного оксида никеля и титана и диодом *p*-CuO/*n*-ZnO : In. Отсутствие обратимого электрического пробоя в диоде не сказывается на работе кроссбара с униполярными мемристорами. Однако в биоморфном нейропроцессоре используются биполярные мемристоры, в которых возможны промежуточные резистивные состояния для реализации синаптических связей между нейронами. Поэтому отработка технологии изготовления комбинированного мемристорно-диодного кроссбара с биполярными мемристорами методом реактивного магнетронного распыления является перспективной.

На рис. 6.12 показана вольт-амперная характеристика диода *p*-Si/$ZnO_x$ (*1*), полученного на «НаноФаб-100» реактивным магнетронным распылением цинка на подложку легированного кремния марки КДБ с удельным сопротивлением 0,03 Ом·см в чистой аргоновой среде [24]. Площадь *p*-*n*-перехода равна 1,0 мм$^2$, толщина пленки $ZnO_x$ 100 нм. Для сравнения приведены ВАХ диодов, полученных электрохимическим осаждением нелегированного ZnO на *p*-кремний с таким же удельным сопротивлением (*2*), спеканием легированных оксидов металлов (*3*) и ионной имплантацией (*4*).

Видно, что характеристики диода, полученного магнетронным способом и спеканием, практически совпадают, при этом, как уже отмечалось, второй способ не пригоден для производства микроэлектроники. Диод, полученный электрохимическим осаждением, имеет худшую характеристику, что связано в первую очередь с малой площадью контакта из-за высокой пористости материала.

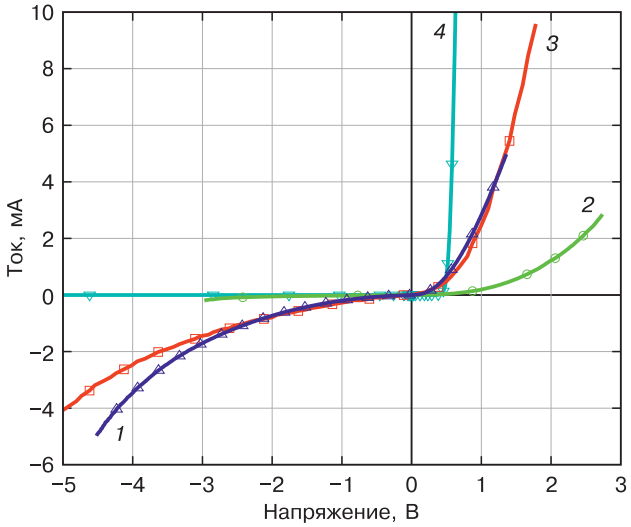



**Рис. 6.12.** Вольт-амперные характеристики диодов, изготовленных по различным технологиям:

*1* — магнетронным осаждением *p*-Si/ZnO;
*2* — электрохимическим осаждением ZnO на *p*-Si; *3* — спеканием слоев CuO/ZnO;
*4* — легирование *p*-Si/*n*-Si ионной имплантацией

Диод, изготовленный с помощью ионной имплантации, имеет пробойное напряжение около −20 В и высокую нелинейность в правой части ВАХ. Но эта технология несовместима с многослойными структурами из-за высокой температуры отжига, как было указано выше.

Изготовление диодных слоев магнетронным распылением с последующим отжигом при 400–600 °C не приводит к разрушению проводников кроссбара. Таким образом, осаждение слоев диода методом магнетронного распыления является наиболее предпочтительным по сравнению с указанными традиционными методами, используемыми в электронике.

### 6.3.2. Выбор материалов полупроводниковых слоев диода с оптимальными характеристиками

Получение нелегированных пленок оксидов металлов с помощью реактивного магнетронного распыления является достаточно простой технологией и заключается в распылении металлического катода в атмосфере кислорода. Внесение примесей в пленку оксида возможно двумя способами: одновременным распылением двух мишеней и использованием уже легированного катода. В [23] приведена вольт-амперная характеристика диода площадью $1{,}37$ нм $\cdot 10^{-11}$ м$^2$ на основе нелегированных оксидов *p*-NiO/*n*-ZnO в узком диапазоне напряжений, в пределах которого не наблюдается обратимого электрического пробоя.

Исследования на магнетронном модуле нанотехнологического комплекса «НаноФаб-100» [24] показали, что вольт-амперная характеристика структуры CuO/ZnO (оба слоя толщиной 45 нм) площадью $1{,}96 \cdot 10^{-11}$ м$^2$ близка к полученной в [17] и не имеет пробоя в широком диапазоне изменения напряжения (рис. 6.13).

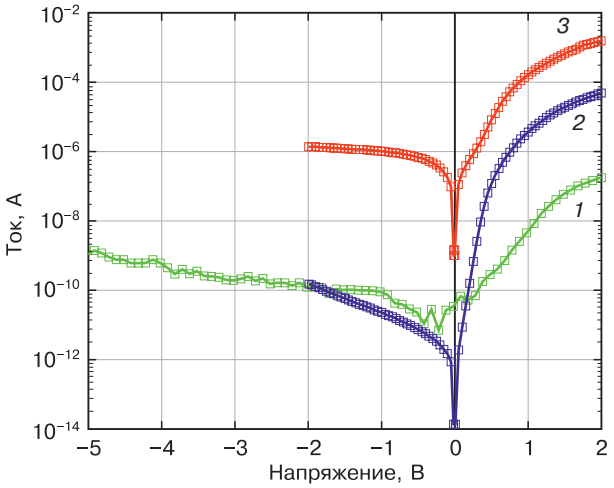



**Рис. 6.13.** Вольт-амперные характеристики диодов из нелегированных оксидов:
*1* — *p*-CuO/*n*-ZnO; *2* — *p*-NiO/*n*-ZnO; *3* — *p*-CuO/*n*-ZnO : In

При использовании легированных оксидов металлов *p*-CuO/*n*-ZnO : In [23] концентрации собственных носителей заряда недостаточно для достижения обратимого электрического пробоя при низких напряжениях (рис. 6.13, *3*). Для решения этой проблемы необходимо один из слоев оксидов металлов заменить на материал с высокой концентрацией носителей заряда, например, легированный кремний.

Для оценки концентрации легированных примесей, обеспечивающей нужное пробойное напряжение, можно воспользоваться формулой [25]:

$$N_{\text{d}} = \varepsilon\varepsilon_0 E^2_{\text{проб}} / 2eV_{\text{проб}},$$

где $E_{\text{проб}}(N_{\text{d}})$ — пробойная напряженность электрического поля в кремнии. Пробойное напряжение диода $V_{\text{проб}} = 2$ В при наличии сильного легирования в одном слое (третья группа: B, Al, Ga) во втором слое соответствует концентрации примесей (пятой группы: N, P, As) $N_{\text{d}} = 1{,}5 \cdot 10^{18}$ см$^{-3}$, что составляет 0,3 % от концентрации атомов кремния. Из элементов пятой группы фосфор и мышьяк невозможно распылять магнетронным способом, а распыление кремниевой мишени в азотной среде не позволяет получить требуемую концентрацию примесей $Si_{99{,}997}N_{0{,}003}$ при минимально возможном уровне натекания азота в магнетроне «НаноФаб-100» 0,1 ссм$^3$/мин.

Альтернативным полупроводником *n*-типа может быть $ZnO_x$, стехиометрию которого можно легко контролировать во время магнетронного распыления. Второй слой диода *p*-типа можно получить распылением легированной бором кремниевой мишени Si : B, поскольку промышленная технология получения кремния разной степени легирования хорошо отработана.

На рис. 6.14 показана зависимость мольной доли цинка в пленке $ZnO_x$ от объемной доли кислорода в общем расходе кислорода и аргона.

Образцы $ZnO_x$ изготовлены с помощью реактивного магнетронного распыления при разных значениях объемной доли $O_2$ в смеси реактивного и рабочего газов $O_2$ и Ar [24]. Измерение мольных долей Zn и O выполнены на сканирующем электронном микроскопе TESCAN MIRA 3 с энергодисперсионным детектором Oxford Instruments Ultim Max и низкой относительной погрешностью (до 2 %).

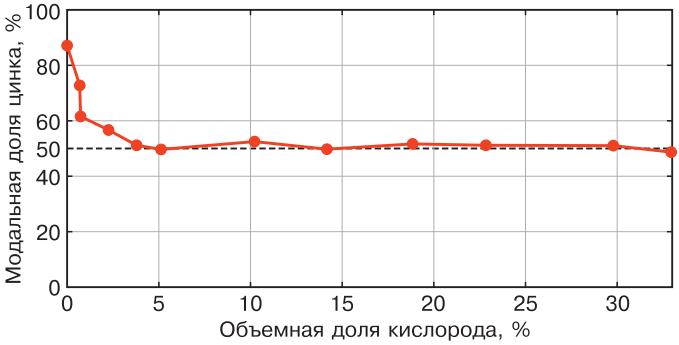



**Рис. 6.14.** Мольная доля цинка в пленке $ZnO_x$ в зависимости от объемной доли кислорода в реактивной газовой смеси

Видно, что в широком диапазоне изменения доли кислорода (от 5 % и выше) состав осаждаемой пленки соответствует чистому оксиду ZnO. Оксид цинка является полупроводником *n*-типа и используется как слой диода (например, [23]). Очевидно, что при увеличении содержания Zn в пленке число электронов проводимости будет увеличиваться, что приводит к уменьшению пробойного напряжения и сопротивления в открытом состоянии.

### 6.3.3. Технология изготовления комбинированного мемристорно-диодного кроссбара

Изготовление комбинированного кроссбара проводилось путем последовательного напыления диодных и мемристорного активных слоев через маски электронного резиста (PMMA) в магнетронном модуле, входящем в нанотехнологический комплекс NT-MDT «НаноФаб-100» [16; 24]. Мемристоры изготовлены по технологии «кроссбар», отработанной в [11]. Экспонирование резиста производилось на электронном микроскопе JSM-6510LV-EDS.

Электромагнитные клапаны на газовых магистралях установки позволяют регулировать расход газов с точностью до 0,1 ссм$^3$/мин. Диаметр распыляемых мишеней — 78 мм. Вакуумная подготовка реакторной камеры осуществлялась при давлении $5 \cdot 10^{-5}$ Па.

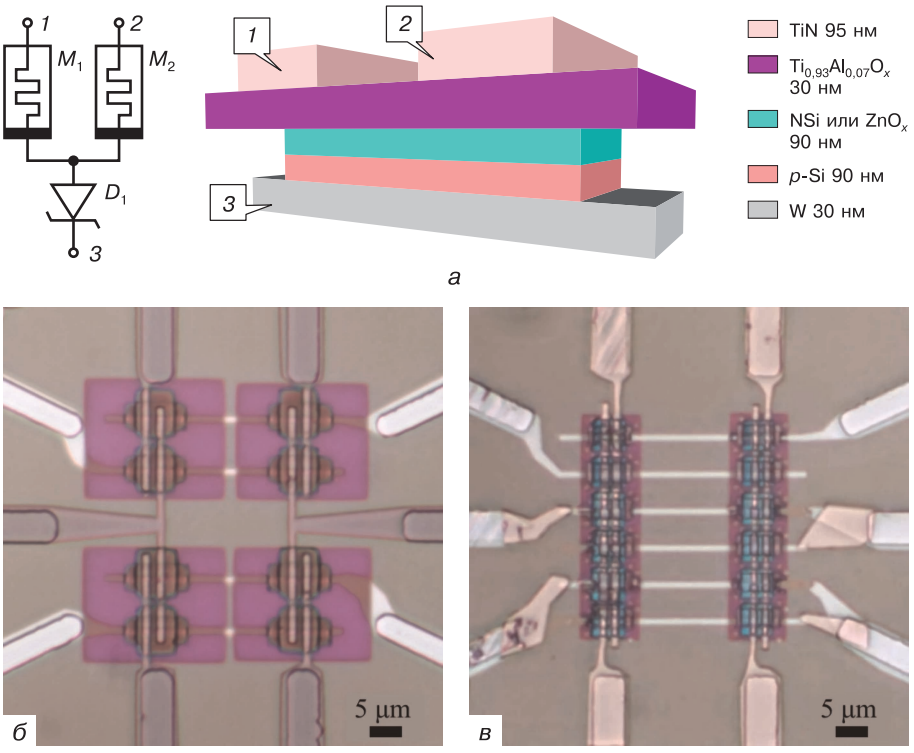



**Рис. 6.15.** Электрическая схема и топология ячейки 1D2M (*а*); микрофотографии изготовленных массивов со структурами: W/*p*-Si/*n*-Si/$Ti_{0,93}Al_{0,07}O_x$/TiN (*б*) и W/*p*-Si/ZnO/$Ti_{0,93}Al_{0,07}O_x$/TiN (*в*)

Сначала на подложку через маску наносились нижние проводники из вольфрама W шириной 1 мкм и толщиной 30 нм с адгезивным 5 нм подслоем Ti. Далее через две другие маски последовательно формировались два полупроводниковых слоя диодов Зенера в результате распыления в магнетроне катода из легированного кремния *p*-Si и реактивного распыления цинка Zn в атмосфере кислорода. Второй вариант диода Зенера со структурой *p*-Si/*n*-Si формировались при последовательном распылении катодов из *p*-Si и *n*-Si кремния. Толщина пленок *p*-Si и $ZnO_x$ (или *n*-Si) равна 90 нм. Затем реактивным магнетронным распылением двух катодов через маски наносилась пленка активного слоя мемристоров толщиной 30 нм на основе смешанного оксида металлов $Ti_{0,93}Al_{0,07}O_x$. Далее на эту пленку через маску наносились верхние проводники кроссбара из нитрида титана TiN

толщиной 95 нм и шириной 1 мкм ортогонально нижним проводникам так, чтобы над каждым диодом проходило два проводника. Таким образом, на кристалле образован один функциональный пласт запоминающей матрицы, состоящий из диодного и мемристорного слоев. Площади мемристоров составили 1 мкм$^2$, а диодов — 15 мкм$^2$. Электрическая схема и топология ячейки 1D2M, а также микрофотографии кроссбар-массивов этих ячеек с двумя разными диодами приведены на рис. 6.15: соответственно, с числом ячеек 2 × 2 (*б*) и 6 × 2 (*в*)

На рис. 6.15 хорошо видны взаимно перпендикулярные проводники кроссбара. Пленка мемристорного слоя $Ti_{0,93}Al_{0,07}O_x$ имеет фиолетовый цвет, верхний слой диода *n*-Si (см. рис. 6.15, *б*), расположенный под мемристорным слоем, имеет бежевый цвет, а верхний слой диода $ZnO_x$ (см. рис. 6.15, *в*) — голубой цвет.

Таким образом, метод магнетронного распыления является оптимальным как для изготовления диодов, так и для мемристоров. Все слои комбинированного мемристорно-диодного кроссбара, включая проводящие дорожки, могут быть изготовлены в одном технологическом цикле.

## 6.4. РАЗРАБОТКА И ИЗГОТОВЛЕНИЕ ИЗМЕРИТЕЛЬНОГО СТЕНДА С УПРАВЛЯЮЩЕЙ ПЕРИФЕРИЙНОЙ ЭЛЕКТРИЧЕСКОЙ СХЕМОЙ

Вначале были разработаны, а затем изготовлены управляющие периферийные электрические схемы на дискретных элементах с КМОП-логикой для обеспечения работы запоминающей и логической матриц, построенных на основе комбинированного мемристорно-диодного кроссбара. Схема измерительного стенда [26], представленная на рис. 6.16, включает в себя: мемристорно-диодный кроссбар с четырьмя ячейками 1D2M, АЦП для измерения напряжения и четырьмя цифровыми портами, подключенными к операционным усилителям для формирования импульсов, выходные преобразователи ток–напряжение на основе операционных усилителей (ОУ). Кроме этого, в эксперименте по ассоциативному самообучению задействуется электрическая схема на основе ОУ, имитирующая частичный функционал нейрона, а именно, сравнение выходного тока кроссбара, преобразованного в напряжение, с пороговым напряжением.

Выходные преобразователи ток–напряжение построены на основе операционных усилителей (ОУ) с резистором 1 МОм в качестве отрицательной обратной связи (1 мкА преобразуется в 1 В). При измерении выходного тока замкнуты переключатели $sw_{00}$, $sw_{01}$, $sw_{10}$ и $sw_{11}$, а при измерении выходного напряжения — $sw_{02}$ и $sw_{12}$. В эксперименте по ассоциативному обучению к одной из выходных линий массива ячеек через переключатель sn дополнительно подключается электрическая схема упрощенного нейрона на основе

ОУ, которая при превышении порога с помощью транзистора шунтирует сопротивления в обратной связи входных операционных усилителей. Это приводит к резкому увеличению коэффициента усиления и, соответственно, большему напряжению импульсов. Конденсаторы в обратной связи ОУ служат для подавления самовозбуждения, возникающего из-за наличия положительной обратной связи через три ОУ и один инвертор на транзисторе.

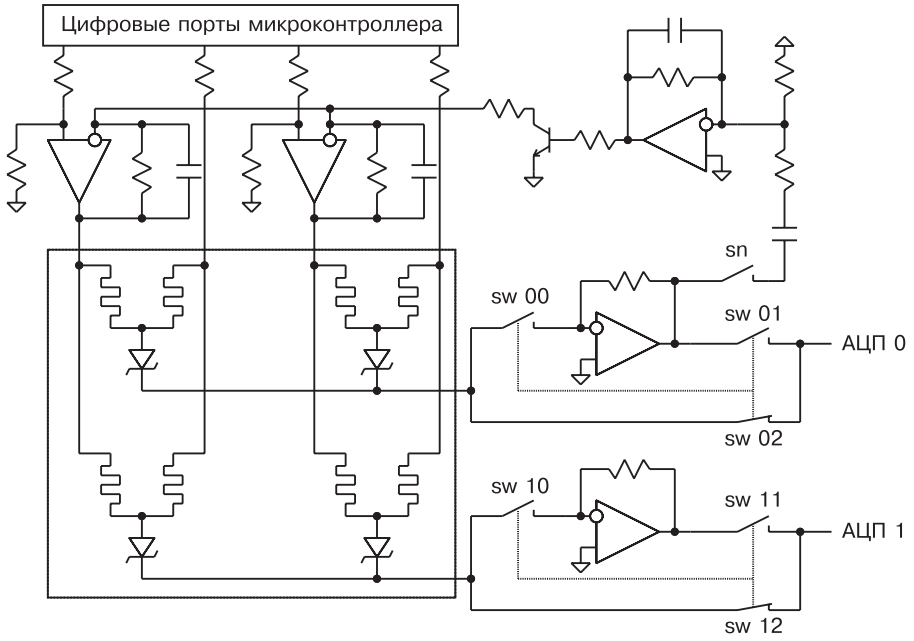



**Рис. 6.16.** Схема измерительного стенда

Измерения электрофизических характеристик лабораторных образцов комбинированного мемристорно-диодного кроссбара и питание стенда осуществляются с помощью программируемого источника-измерителя Keithley SourceMeter 2400.

## 6.5. ИССЛЕДОВАНИЕ ЭЛЕКТРИЧЕСКИХ СВОЙСТВ МЕМРИСТОРНО-ДИОДНОГО КРОССБАРА

### 6.5.1. Электрические свойства мемристора

С помощью лабораторного тестера — одноканального программируемого источника-измерителя Keithley SourceMeter 2400 проведены измерения электрических характеристик мемристора $W/Ti_xAl_{1-x}O_y/TiN$ [11] с активным слоем из чистого оксида титана и с различным содержанием Al. На рис. 6.17, *а* красным цветом показана вольт-амперная характеристика

мемристорного устройства на основе слоя из диоксида титана. Отношение сопротивлений в низкопроводящем и высокопроводящем состояниях при напряжении считывания 0,2 В имеет величину $R = R_{off}/R_{on} \approx 1{,}3$. Для сравнения, в [27] это отношение в мемристоре со слоем диоксида титана толщиной 30 нм и платиновыми электродами равняется двум. Это напряжение выбрано в линейной области вольт-амперной характеристики, которое используется при работе мемристоров в аналоговом режиме.

Зеленым цветом на рис. 6.17, *а* показана вольт-амперная зависимость в мемристоре с активным слоем $Ti_{0,93}Al_{0,07}O_x$. Внесение примеси Al в $TiO_2$ на уровне 7 ат. % увеличивает $R$ с 1,3 до 7,2. Дальнейшее увеличение доли примеси Al не приводит к росту отношения $R$. Максимальное отношение сопротивлений $R$ в мемристоре с активным слоем $Ti_{0,9}Al_{0,1}O_x$ при напряжении считывания 0,2 В равно 3,78 и с активным слоем $Ti_{0,85}Al_{0,15}O_x$ соответственно 2,41.

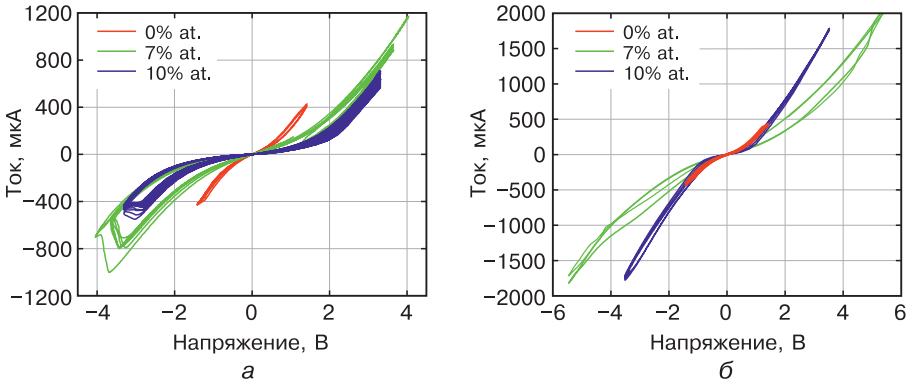



**Рис. 6.17.** ВАХ структур:
*а* — $W/Ti_xAl_{1-x}O_y/TiN$ [6]; *б* — $W/Ti_xZr_{1-x}O_y/TiN$ при разной мольной доле примеси

Исследование мемристорной структуры $W/Ti_xAl_{1-x}O_y/TiN$ показало (рис. 6.18) существование оптимальной мольной доли примеси Al, равной 7 % ат., при которой достигается максимальное отношение сопротивлений мемристора в низкопроводящем и высокопроводящем состояниях равное 7,2. Площадь мемристора составила 1 мкм$^2$, толщина активного слоя 30 нм.

Пленка смешанного оксида $Ti_xZr_{1-x}O_y$ получена аналогичным образом в результате одновременного распыления двух катодов Ti и Zr в атмосфере аргона и кислорода [26]. Толщина активного слоя также составила 30 нм, а площадь 1 мкм$^2$. На рис. 6.17, *б* приведены ВАХ структуры $W/Ti_xZr_{1-x}O_y/TiN$ при разной доле Zr в активном слое оксида титана. Откуда следует, что в этой структуре также существует оптимальная мольная доля примеси, близкая к 7 % ат., при которой достигается максимальное отношение сопротивлений мемристора в низкопроводящем и высокопроводящем состояниях, равное 1,6 при 2 В.

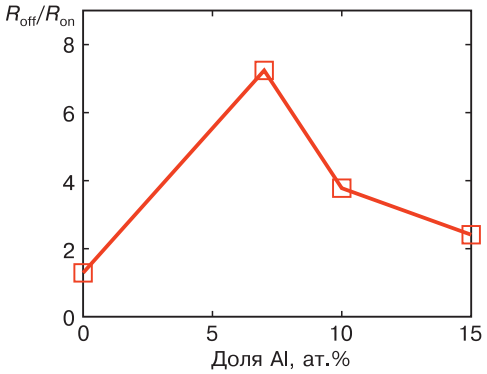



**Рис. 6.18.** Зависимость отношения сопротивлений в низкопроводящем $R_{off}$ и высокопроводящем $R_{on}$ состояниях при напряжении считывания 0,2 В от доли примеси Al

Таким образом, из сравнения вольт-амперных характеристик следует, что предпочтительным материалом активного слоя мемристора является $Ti_{0,93}Al_{0,07}O_y$. Повысить отношение сопротивлений $R$ в этом мемристоре можно, уменьшив долю кислорода. В [2] показано, что обедненный кислородом $TiO_x$ обладает значением $R = 100$, соизмеримым с чистыми оксидами Zn, Zr и Hf.

Можно ожидать, что в мемристорах на смешанных оксидах $Ti_xSc_{1-x}O_y$, $Hf_xSc_{1-x}O_y$, $Hf_xY_{1-x}O_y$, $Hf_xLu_{1-x}O_y$, $Zr_xSc_{1-x}O_y$, $Zr_xY_{1-x}O_y$, $Zr_xLu_{1-x}O_y$, также будет наблюдаться оптимальная доля примеси, соответствующая максимально повышенному отношению сопротивлений в низкопроводящем и высокопроводящем состояниях.

### 6.5.2. Электрические свойства диода Зенера

На рис. 6.19, *а* показаны вольт-амперные характеристики диода площадью $9{,}7 \cdot 10^3$ мкм$^2$ на основе гетероперехода $p$-Si/$ZnO_x$ при разных мольных долях Zn и O [26]. Образцы $ZnO_x$ изготовлены с помощью реактивного магнетронного распыления при разных значениях объемной доли $O_2$ в смеси реактивного и рабочего газов $O_2$ и Ar. Измерение мольных долей Zn и O выполнены на сканирующем электронном микроскопе TESCAN MIRA 3 с энергодисперсионным детектором Oxford Instruments UltimMax и низкой относительной погрешностью (до 2 %).

Из рис. 6.19, *а* видно, что при прямом включении гетероперехода ток возрастает экспоненциально, а при обратном включении происходит обратимый пробой при малом напряжении. Если отразить вольт-амперные характеристики относительно начала координат (включить диод в обратном направлении), то кривые будут повторять ход вольт-амперных характеристик диодов на рис. 6.12. Кроме этого, с увеличением мольной доли цинка нелинейность ВАХ растет, достигает максимума при 61,75 ат. %, а затем

падает. Увеличение доли цинка приводит к росту числа основных носителей заряда (электронов), что увеличивает нелинейность. Однако слишком большая концентрация цинка приводит к шунтированию *p*-*n*-перехода, что в свою очередь приводит к выравниванию вольт-амперной характеристики. Таким образом, существует оптимальная мольная доля Zn, которая дает наилучшие характеристики диода *p*-Si/ZnO.

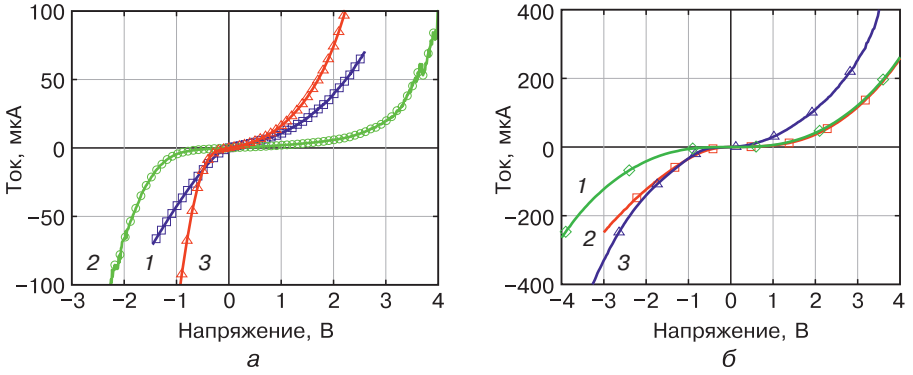



**Рис. 6.19.** ВАХ диода:

*а* — на основе гетероперехода *p*-Si/$ZnO_x$ при разной мольной доле Zn: *1* — 56,68 %, *2* — 61,75 %, *3* — 72,75 % с удельным сопротивлением *p*-Si, равным 0,275 Ом · см [24]; *б* — на основе гетеропереходов [26] *p*-Si/*n*-Si (*1*) и *p*-Si/$Zn_{0,62}O_{0,38}$ с разным удельным сопротивлением *p*-Si: *2* — 0,066 Ом · см и *3* — 0,275 Ом · см

Дальнейшее улучшение вольт-амперной характеристики диода достигается за счет увеличения уровня легирования *p*-Si-слоя. На рис. 6.18, *б* представлены вольт-амперные характеристики диода площадью 1 мкм$^2$ на основе гетероперехода *p*-Si/$ZnO_x$, в котором слои *p*-Si имеют разный уровень легирования бором (разное удельное сопротивление). Видно, что с увеличением уровня легирования слоя *p*-Si происходит рост нелинейности вольт-амперной характеристикой и увеличение напряжения обратимого пробоя. Поскольку напряжение обратимого пробоя меньше напряжения открытия диода при прямом включении гетероперехода полученный диод в электрической схеме ячейки кроссбара должен быть включен в обратном направлении. Меньший уровень легирования приводит к более широкому интервалу напряжений, при котором диод остается закрытым. При этом кривая *2* отражает большее напряжение открытия при прямом включении и меньшее пробойное напряжение (–0,2 В против –0,6 В на кривой *1*), что ближе к сформулированным требованиям для селективного элемента кроссбара. Низкое напряжение пробоя изготовленного диода обусловлено сильной степенью легирования полупроводниковых слоев. Область тока утечки для структуры *p*-Si/*n*-Si находится в диапазоне (–0,6 В, 0,6 В), а для структуры *p*-Si/$ZnO_x$ (кривая *2*) — в диапазоне (–0,2 В, 0,6 В).

Вольт-амперные характеристики диода $p$-Si/$ZnO_x$, представленные на рис. 6.18, *б*, имеют слабую нелинейность по сравнению с высокими характеристиками кремниевого диода на переходе $p$-Si/$n$-Si, изготовленного магнетронным методом. Отсюда сделан вывод, что структура $p$-Si/$n$-Si является наиболее приемлемой с точки зрения применения ее в качестве диода в комбинированном мемристорно-диодном кроссбаре, поскольку она лучше удовлетворяет указанным требованиям к селективному элементу, а также хорошо вписывается в кремниевую технологию создания элементов микроэлектроники.

### 6.5.3. Электрические свойства мемристорно-диодной ячейки

Вольт-амперные характеристики ячеек мемристорно-диодных кроссбаров TiN/$Ti_{0,93}Al_{0,07}O_x$/$p$-Si/$n$-Si/W TiN/$Ti_{0,93}Al_{0,07}O_x$/$p$-Si/ZnO/W [26] представлены на рис. 6.20. На зеленой кривой рис. 6.20, *б* резкого перехода RESET не видно из-за принудительного ограничения максимального тока. Отличие в ширине гистерезиса на рис. 6.17, *а* и рис. 6.20, *б* обусловлено различием материала одного из электродов мемристора. В первом случае материалом электродов мемристора является TiN. а во втором — TiN и $n$-Si.

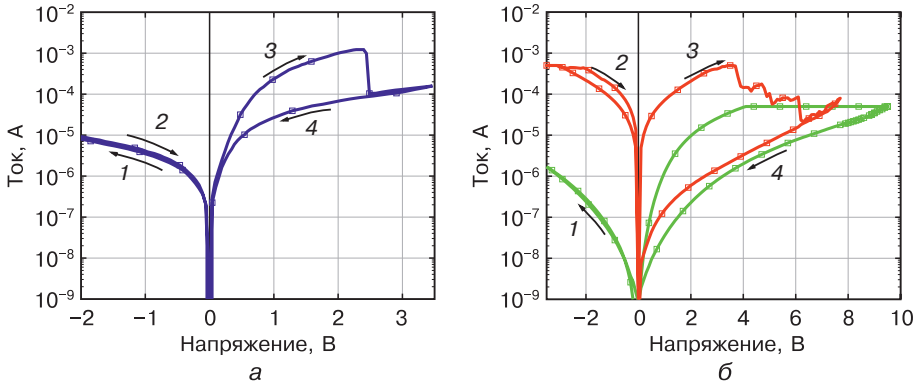



**Рис. 6.20.** Вольт-амперные характеристики:
*а* — ячейки кроссбара из работы [23]; *б* — красным цветом ячейка кроссбара с диодом $p$-Si/$ZnO_x$ и зеленым цветом — с диодом $p$-Si/$n$-Si

Большое сопротивление закрытого диода приводит к стягиванию гистерезиса в обратной ветви вольт-амперной характеристики ячейки, поскольку вклад сопротивления диода преобладает над вкладом малого сопротивления мемристора в их общей ВАХ. Такой же эффект наблюдается в кроссбаре [23] (рис. 6.20, *а*) с униполярным мемристором на основе смешанного оксида никеля и титана и диодом $p$-CuO/$n$-ZnO : In.

Как видно из рис. 6.12, *б* ячейка с диодом *p*-Si/*n*-Si обладает лучшим выпрямляющим свойством по сравнению с ячейкой с диодом *p*-Si/$ZnO_x$, поскольку ток в открытой ячейке при положительном напряжении значительно выше, чем при отрицательном напряжении. Высокое выпрямляющее свойство ячейки необходимо для функционирования диодной логики в логической матрице и при записи состояний мемристоров в запоминающей и логической матрицах.

## 6.6. ИССЛЕДОВАНИЕ ПРОЦЕССОВ ОБРАБОТКИ СИГНАЛОВ В КРОССБАРАХ ДЛЯ ЗАПОМИНАЮЩЕЙ И ЛОГИЧЕСКОЙ МАТРИЦ

### 6.6.1. Сложение выходных импульсов нейронов

Входной импульс подается на контакты выбранной ячейки 1D2M кроссбара запоминающей матрицы в виде двух импульсов напряжения противоположной полярности, абсолютная величина которых меньше порогового напряжения переключения мемристора. В результате комплементарная пара мемристоров образует резистивный делитель напряжения. Выходное напряжение ячейки однозначно определяется соотношением сопротивлений мемристоров. Выходные напряжения ячеек складываются при подключении выходной линии кроссбара к усилителю с высокоомным входом и нагрузкой в виде конденсатора. Сложение токов, протекающих через закрытые диоды Зенера, происходит при низком входном сопротивлении усилителя. Взвешивание напряжений входных импульсов с последующим суммированием токов уменьшает паразитные токи между ячейками.

На рис. 6.21 показан выходной ток одной шины матрицы (см. рис. 6.15) как результат сложения токов из двух ячеек, находящихся в разных синаптических состояниях [28]. Представлено три случая: в первом входной импульс напряжения подается только на первую ячейку, во втором — только на вторую, а в третьем — на обе. При этом первая ячейка имеет положительный вес, вторая — отрицательный. Амплитуда выходного тока в третьем случае должна быть равна сумме амплитуд токов в первом и во втором случаях, однако она оказывается меньше ожидаемой, что может быть связано с влиянием нелинейной ВАХ селективного диода.

В качестве модели ячеек при SPICE-моделировании использована экспериментально полученная вольт-амперная зависимость, представленная на рис. 6.20, *б*. Результат моделирования показывает, как выполняется суммирование токов на выходной шине кроссбара при отсутствии разброса характеристик мемристоров. Среднеквадратичное отклонение тока за время действия входных импульсов напряжения составляет 234 нА при среднеквадратичном отклонении сопротивления мемристоров в высокопроводящем и низкопроводящем состояниях 65 и 109 % соответственно.

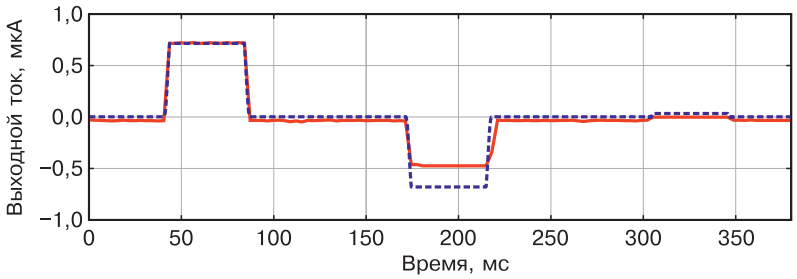



**Рис. 6.21.** Сложение выходных токов двух ячеек 1D2M с комплементарными мемристорами на общей выходной шине кроссбара (см. рис. 6.15, *б*):

*сплошная кривая* — экспериментальная;
*пунктирная кривая* — SPICE-моделирование

### 6.6.2. Маршрутизация импульсов на синапсы других нейронов

Логическая матрица с ячейками 1D1M работает в качестве маршрутизатора, направляя выходные импульсы нейронов из нейронного блока на синапсы других нейронов в запоминающей матрице. При этом в кроссбаре реализуются логические вентили «ИЛИ» с помощью диодно-резисторной логики на основе селективных диодов и резисторов, подтягивающих выходные проводники кроссбара к низкому электрическому потенциалу. На рис. 6.22 показаны выходные напряжения матрицы для трех случаев маршрутизации и их представление в виде логического умножения булевой матрицы (матрицы преобразования) на входной булевый вектор [28]. Первая компонента выходного вектора обозначена как Out [0], вторая — Out [1].

В каждом из трех случаев сначала импульс напряжения появляется на первом входе маршрутизатора, а затем на втором. Матрица преобразования первого случая обеспечивает только перенаправление импульса напряжения с первого входа на второй выход. Матрица преобразования во втором случае обеспечивает взаимное перенаправление импульсов: первый импульс появляется на втором выходе, а второй импульс — на первом. В третьем случае первый импульс пройдет на оба выхода, а второй импульс — только на первый.

Экспериментально полученные амплитуды напряжения выходных импульсов накладывают ограничения на порог переключения инверторов логической матрицы. Порог напряжения должен быть выбран таким образом, чтобы выходной вектор после восстановления логических уровней нуля и единицы при прохождении через инвертор соответствовал выходному вектору, полученному по математическим правилам. В данном случае логическая матрица будет работоспособна при пороговом напряжении, равном 0,5 В.

Среднеквадратичное отклонение выходных напряжений за время действия входных импульсов напряжения, вносимая неодинаковостью мемристоров, оценена при сравнении данных эксперимента и SPICE-моделирования и составляет 668 мкВ. Среднеквадратичное отклонение сопротивления мемристоров в низкопроводящем состоянии составляет 137 %, а в высокопроводящем — 97 %.

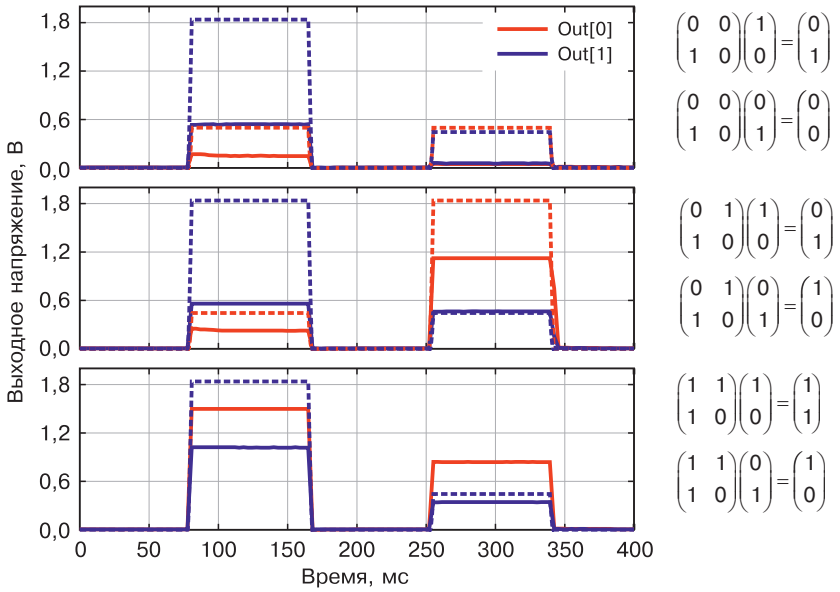



**Рис. 6.22.** Маршрутизация импульсов
с помощью запрограммированных мемристоров:
*сплошные кривые* — экспериментальные;
*пунктирные кривые* — SPICE-моделирование

### 6.6.3. Умножение матрицы чисел на вектор

Запоминающая матрица представляет собой массив синапсов аппаратной нейронной сети. На входные проводники кроссбара запоминающей матрицы приходят импульсы от нейронов, амплитуда напряжения которых затем умножается на весовые коэффициенты, определяющиеся состоянием мемристоров в ячейках матрицы. Получившиеся токи складываются на выходных проводниках как описано в подразд. 6.5.1.

Наличие импульсов на входных проводниках в определенный момент времени можно описать с помощью вектора. Поскольку амплитуда импульсов от нейронов всегда одна и та же, то отсутствие импульса можно описать нулем, а наличие — единицей. Состояния ячеек запоминающей матрицы при этом можно описать матрицей весовых коэффициентов. Тогда, в любой

момент времени выходные токи запоминающей матрицы будут определяться вектором, являющимся результатом умножения матрицы весовых коэффициентов на вектор входных напряжений.

На рис. 6.23 представлен результат работы запоминающей матрицы размером 2 × 2 для трех разных матриц весовых коэффициентов [28]. Первая компонента вектора выходных токов обозначена как Out [0], вторая — Out [1].

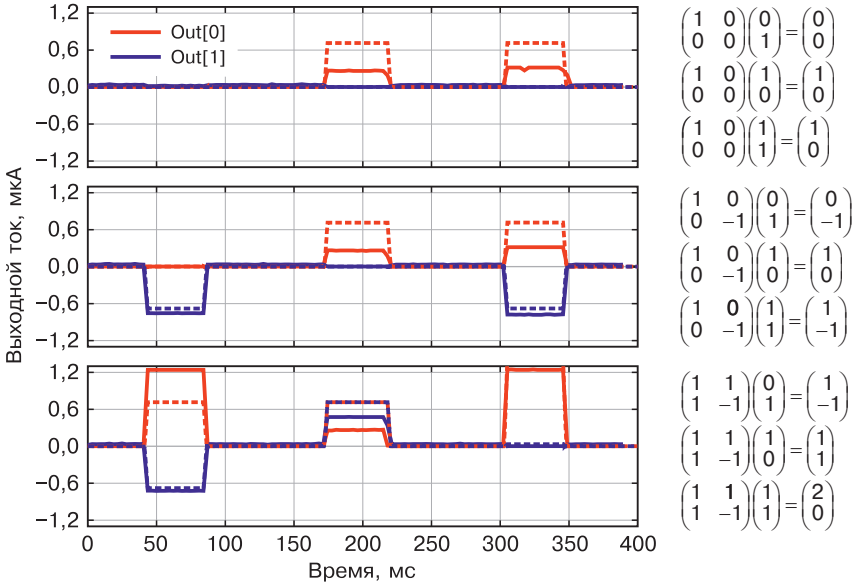



**Рис. 6.23.** Результат умножений матрицы чисел размером 2 × 2 на двухкомпонентный вектор:

*сплошные кривые* — экспериментальные;
*пунктирные кривые* — SPICE-моделирование

Как следует из рис. 6.23, выходной вектор соответствует ожидаемому результату, вычисленному по правилу матрично-векторного умножения. Отсутствие ожидаемого удвоения амплитуды выходного тока на первом выходе запоминающей матрицы при выполнении последнего умножения вызвано высоким коэффициентом усиления преобразователя ток-напряжение, что привело к превышению максимального напряжения оцифровки АЦП-микроконтроллера.

Среднеквадратичное отклонение выходных напряжений за время действия входных импульсов напряжения, вносимая неодинаковостью мемристоров, оценена при сравнении данных эксперимента и SPICE-моделирования и составляет 276 нА. Среднеквадратичное отклонение сопротивления мемристоров в низкопроводящем состоянии составляет 109 %, а в высокопроводящем — 65 %. При меньших отклонениях сопротивле-

ний мемристоров в запоминающей матрице электрическая схема, выполняющая умножение матрицы на вектор, остается работоспособной.

Из результатов моделирования обработки единичного импульса в запоминающей матрице [29] следует, что существенная деградация напряжения выходного импульса (с 1 до 0,3 В) происходит в матрице размером 1000 × 1000 ячеек при отношении максимального и минимального сопротивлений мемристора, равном 100. Увеличивать размер изготавливаемой работоспособной матрицы можно путем повышения отношения максимального и минимального сопротивления мемристора, а также путем уменьшения разброса этих характеристик.

### 6.6.4. Паразитные токи в соседних мемристорно-диодных ячейках

Организация ячеек в кроссбаре без селективных элементов обеспечивает большую плотность элементов, но обладает существенным недостатком, связанным с протеканием паразитных токов через соседние ячейки. Обработка сигналов при минимизации паразитных токов в соседних мемристорных ячейках с диодом Зенера необходима для достижения устойчивого режима и энергоэффективной работы сверхбольших матриц.

На рис. 6.24, *а* показано изменение выходного тока ячейки 1D2M изготовленного кроссбара для запоминающей матрицы с одним мемристором в высокопроводящем состоянии в зависимости от числа соседних ячеек в таком же состоянии (рис. 6.24, *б*). Это изменение выходного тока связано с протеканием паразитных токов через соседние ячейки. Таким образом проявляется взаимовлияние ячеек. Изначально мемристоры в соседних ячейках находятся в низкопроводящем состоянии.

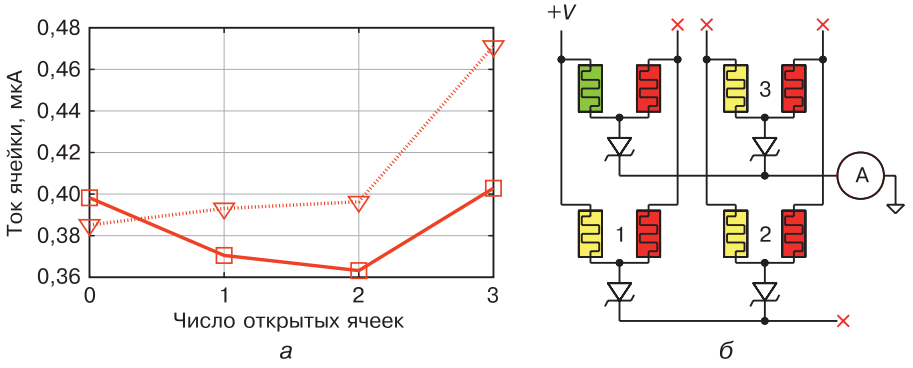



**Рис. 6.24.** Изменения выходного тока ячейки запоминающей с открытыми мемристорами в зависимости от числа соседних открытых ячеек:

*сплошная кривая* — экспериментальная;
*пунктирная кривая* — моделирование

Как видно на рис. 6.24, SPICE-моделирование с одинаковыми усредненными ячейками показывает монотонный рост тока измеряемой ячейки при увеличении числа соседних ячеек с высокопроводящими мемристорами. Но в эксперименте уменьшение сопротивления мемристоров в соседних ячейках не всегда приводит к наблюдаемому увеличению тока через измеряемую ячейку.

Это можно объяснить следующим образом. При переключении мемристоров в соседних ячейках в них происходит изменение концентрации кислорода вблизи электродов, что приводит к изменению контактной разности потенциалов. И, как следствие, ток через измеряемую ячейку уменьшается.

Управляя параметрами диода Зенера, можно уменьшить энергопотребление при работе комбинированного кроссбара. Из табл. 6.2, в которой приведены параметры диодов, следует, что при увеличении нелинейности вольт-амперной характеристики диода снижается энергопотребление кроссбара как в запоминающей, так и в логической матрицах.

*Таблица 6.2*

**Усредненное энергопотребление кроссбара 2 × 2 в логической и запоминающей матрицах**

| Кроссбар с диодом | Сопротивление диода при прямом смещении | Сопротивление диода при обратном смещении | Энерго-потребление в запоминающей матрице | Энерго-потребление в логической матрице |
|---|---|---|---|---|
| *p*-Si/ZnO | 97,3 кОм (1,5 В)<br>25,8 кОм (3 В) | 20,1 кОм (1,5 В)<br>12,1 кОм (3 В) | 1,41 мкВт | 6,46 мкВт |
| *p*-Si/*n*-Si | 80,7 кОм (1,5 В)<br>24,5 кОм (3 В) | 80,9 кОм (1,5 В)<br>24,7 кОм (3 В) | 1,26 мкВт | 4,58 мкВт |

Энергопотребление матриц найдено с помощью SPICE-моделирования экспериментов, приведенных на рис. 6.22 и рис. 6.23, путем расчета средней суммы мощностей источников входных напряжений.

### 6.6.5. Демонстрация принципа ассоциативного самообучения синапсов запоминающей матрицы

Ассоциативное самообучение запоминающей матрицы происходит, когда информационные импульсы, формируемые выходом возбужденного нейрона в нейросети, приводят к усилению его синаптических связей. Усиление связей возбужденного нейрона с нейронами предыдущего слоя имеет место, если в этот момент времени эти нейроны тоже оказались возбужденными

и сформировали на своих выходах информационные импульсы. Обучение синапсов в запоминающей матрице происходит по правилу Хебба, как и в реальном синапсе: сила связи между одновременно активировавшимися нейронами увеличивается. Для реализации индуцированной долговременной потенциации синапса была выбрана схема из трех нейронов. Выходные импульсы двух виртуальных нейронов через синаптические связи, представленные комплементарными мемристорно-диодными ячейками, поступают на вход третьего реального нейрона. При этом один синапс является сильным (с высоким весовым коэффициентом), второй — слабым (с низким).

Каждый раунд обучения состоит из двух шагов. На первом шаге входные импульсы присутствуют на оба синапса нейрона, на втором шаге импульс приходит только на слабый синапс. Проходя через сильный синапс, импульс вызывает активацию нейрона, что в контексте измерительного стенда выражается в превышении порогового напряжения и последующего увеличения амплитуды входных импульсов посредством изменения коэффициента обратной связи входных ОУ. При этом напряжение на одном мемристоре ячейки слабого синапса становится больше порогового. Процесс обучения слабого синапса в течение нескольких раундов обучения представлен на рис. 6.25 [28].


**Рис. 6.25.** Импульсы напряжения, отвечающие импульсам выходного тока одной из линий запоминающей матрицы, полученные в ходе 6 раундов ассоциативного самообучения. Генерация новой ассоциации

В ходе третьего раунда обучения сопротивление мемристора ячейки слабого синапса становится достаточно малым, чтобы вызвать открытие диода в ячейке и появление наблюдаемого импульса тока. При последующем переобучении нейросеть опирается на ассоциации, заложенные искусственно. Увеличение весового коэффициента изначально слабого синапса отражает возникновение новой ассоциативной связи в сети из трех нейронов.

## 6.7. ИЗГОТОВЛЕНИЕ И ТЕСТИРОВАНИЕ АППАРАТНОЙ НЕЙРОСЕТИ ПРОЦЕССОРА

В [30] представлены результаты по исследованию работы аппаратной импульсной нейросети с мемристорными синапсами в виде запоминающей матрицы в режиме расчета синапсов однослойного персептрона. Персептрон может рассматриваться в качестве первого слоя биоморфной нейросети [31], выполняющего первичную обработку поступающей информации в биоморфном нейропроцессоре.

### 6.7.1. Электрическая схема аппаратного персептрона на основе запоминающей матрицы с мемристорными синапсами

Аппаратный персептрон построен на основе мемристорно-диодного кроссбара с четырьмя парами входных проводников и двумя выходными шинами. Соответственно, кроссбар содержит восемь ячеек, являющихся синапсами нейросети. Слой персептрона образован двумя нейронами, построенными на основе операционных усилителей. Электрическая схема нейрона состоит из преобразователя ток–напряжение, аналогового интегратора, компаратора, схемы задержки в виде интегрирующей RC-цепи и полевого транзистора (рис. 6.26).

Преобразователь ток–напряжение, являющийся входом нейрона, поддерживает на выходных шинах кроссбара виртуальный нулевой потенциал, обеспечивая сложение выходных токов синапсов. Пропорциональное входному синаптическому току напряжение поступает на интегратор, имитирующий накопление заряда на мембране нейрона.

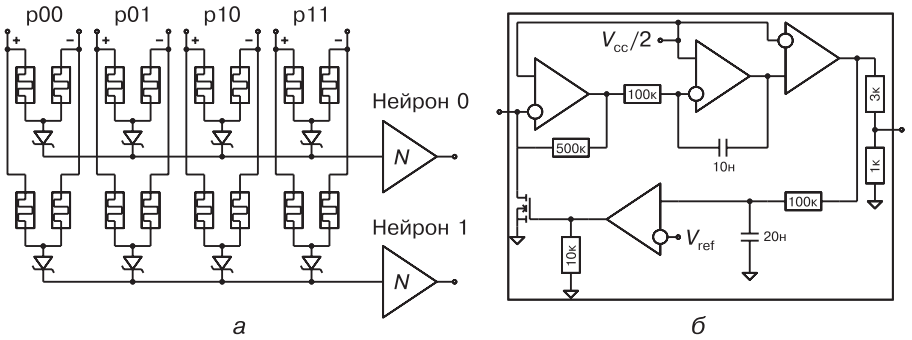



**Рис. 6.26.** Реализация аппаратной нейронной сети на основе мемристорно-диодного кроссбара:
*а* — включение кроссбара в качестве массива синапсов;
*б* — электрическая схема нейрона

Выходное напряжение интегратора сравнивается с пороговым на компараторе. При превышении порога происходит переключение компаратора, вызывающее открытие полевого транзистора, что в свою очередь приводит к разряду конденсатора в интеграторе и появлению на соответствующей выходной шине кроссбара отрицательного потенциала по отношению к виртуальному нулевому потенциалу. Если входной импульс нейросети в данный момент времени будет положительным, произойдет изменение сопротивления одного из мемристоров ячейки, поскольку падение напряжения на нем будет выше порога переключения мемристора.

С помощью магнетронной технологии [24; 28] был изтовлен мемристорно-диодный кроссбар запоминающей матрицы с числом ячеек 4 × 2, представляющий собой интегральный массив синапсов аппаратного персептрона (рис. 6.27).

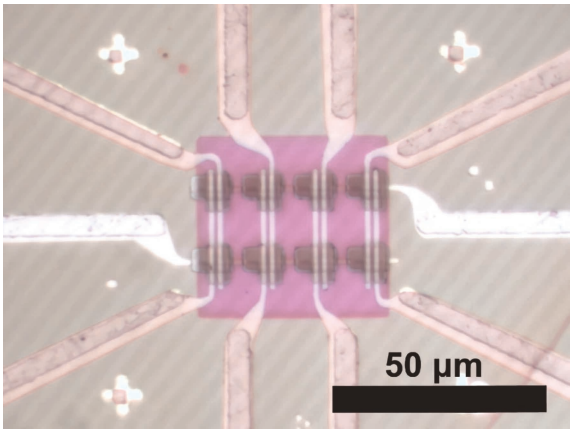



**Рис. 6.27.** Микрофотография мемристорно-диодного кроссбара $TiN/Ti_{0,93}Al_{0,07}O_x/p$-Si/$n$-Si/W

Исследуемая аппаратная импульсная нейросеть на основе мемристорно-диодного кроссбара отличается от [28] большим числом нейронов и синапсов.

### 6.7.2. Универсальный стенд для исследования аппаратной импульсной нейросети

Для исследования работы аппаратной импульсной нейросети разработан и изготовлен измерительный стенд, электрическая схема которого состоит из схемы аппаратного персептрона и входной периферийной электрической схемы для реализации активационной функции нейрона и обеспечения работы запоминающей матрицы в импульсном режиме (рис. 6.28).

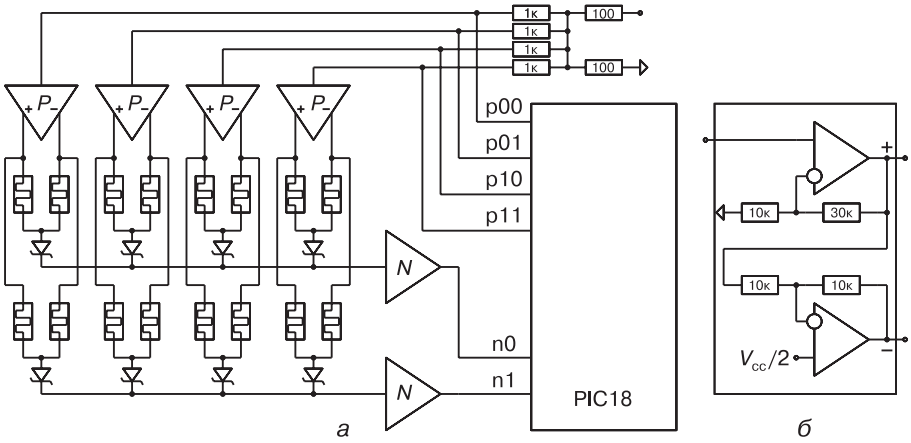



**Рис. 6.28.** Электрические схемы:
*а* — стенда с аппаратной нейросетью;
*б* — формирователя импульсов противоположных полярностей

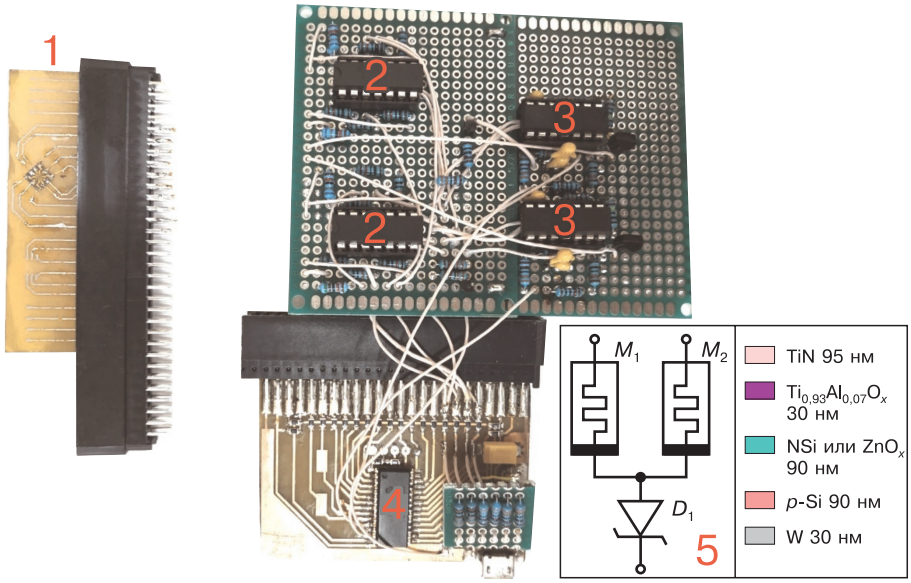



**Рис. 6.29.** Измерительный стенд для тестирования аппаратной импульсной нейросети

Электрическая схема измерительного стенда для тестирования аппаратной импульсной нейросети реализована в виде трех отдельных плат (рис. 6.29) и состоит из схемы аппаратного персептрона и периферийной управляющей электрической схемы (4). Изготовленный кроссбар (1) с топологией (5) и распаянный на отдельной плате, вставляется в разъем,

который позволяет исследовать работу нейросети с разными мемристорными кроссбарами без переделки стенда. Формирователь входных импульсов четырех виртуальных нейронов (2) построен на основе операционных усилителей (ОУ) и обеспечивает усиление импульсов от управляющего микроконтроллера и создание инверсных по напряжению импульсов.

Электрические схемы двух выходных аппаратных нейронов (3) также построены с применением ОУ. Преобразователь ток–напряжение, являющийся входом нейрона, поддерживает на выходных шинах кроссбара виртуальный нулевой потенциал, обеспечивая сложение выходных токов синапсов. Коэффициент преобразования определялся с помощью ВАХ ячеек.

Измерительный стенд [26] был предназначен для обеспечения работы запоминающей матрицы на основе комбинированного мемристорного-диодного кроссбара. В [26] изменение состояния мемристоров обеспечивалось изменением амплитуды входных импульсов и обучение нейросети производилось по механизму долговременной синаптической потенциации (LTP — long term potentiation). Правило LTP применимо в узком числе задач, поскольку всегда приводит к усилению связей, что может негативно сказаться на работе нейросети. Реализованный в новом стенде механизм пластичности, зависимой от времени импульса (STDP — spike time dependent plasticity) учитывает причинность. Если пресинаптический импульс пришел раньше и перекрывается во времени с постсинаптическим, то есть он может быть причиной активации постсинаптического нейрона и вес синапса возрастает. Если же постсинаптический импульс возник раньше пресинаптического, то он не может быть причиной активации постсинаптического нейрона, и вес синапса уменьшается.

Специальная форма входных импульсов позволяет реализовать ассоциативное обучение нейросети по механизму STDP. Изменение веса синапса зависит от разницы между временами срабатывания пресинаптического и постсинаптического нейронов $\Delta t = t_{\text{pre}} - t_{\text{post}}$. Форма пре- и постсинаптических импульсов для обеспечения механизма STDP и соответствующая функция пластичности представлены на рис. 6.30. Величина $\Delta Q$ показывает изменение заряда конденсатора в интеграторе нейрона при прохождении одного информационного импульса после изменения веса синапса.

Так как электрическая схема нейрона выполняет интегрирование синаптического тока во времени, форма входных импульсов должна быть асимметричной. Площадь области действия положительного напряжения должна быть больше площади действия отрицательного напряжения, чтобы обеспечить накопление заряда на конденсаторе интегратора. В свою очередь эта несимметричность ведет к несимметричной функции пластичности с преобладанием области усиления веса синапса. Увеличение весового коэффициента синапса (левая область функции пластичности на рис. 6.30, *б*) соответствует возникновению новой ассоциации (нового знания) в нейросети.

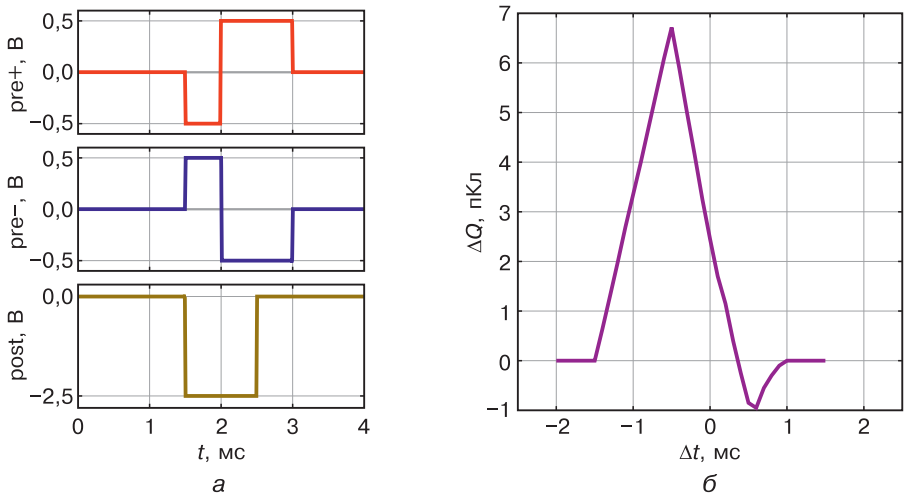



**Рис. 6.30.** Пластичность, зависимая от времени импульса:
*а* — пре- и постсинаптические напряжения ячейки запоминающей матрицы при $\Delta t = 0$;
*б* — функция пластичности $\Delta Q(\Delta t)$

### 6.7.3. Экспериментальное исследование электрических свойств ячеек кроссбара

Для настройки стенда необходимо знать электрические характеристики мемристорно-диодного кроссбара. Амплитуда входных импульсов для кроссбара относительно электрического потенциала виртуального нуля должна быть меньше напряжения переключения мемристоров, но обеспечивать превышение этого порога при открытии полевого транзистора нейрона.

Измеренная вольт-амперная характеристика (ВАХ) ячейки кроссбара при подаче напряжения на один из мемристоров пары показана на рис. 6.31, *а*.

Сопротивление ячейки в закрытом состоянии, измеренное при напряжении входных импульсов определяет коэффициент усиления преобразователя ток–напряжение в электрической схеме нейрона равным 500 В/А. Усиление должно быть достаточно для срабатывания нейрона в начале обучения нейросети.

Из рис. 6.31, *б* видно, что до 6 В изменения сопротивления практически не происходит. Поэтому информационные импульсы не должны превышать этот порог. Меньшее наблюдаемое значение порогового напряжения на рис. 5, *а* связан с гораздо более медленным изменением напряжения на мемристоре при снятии ВАХ.

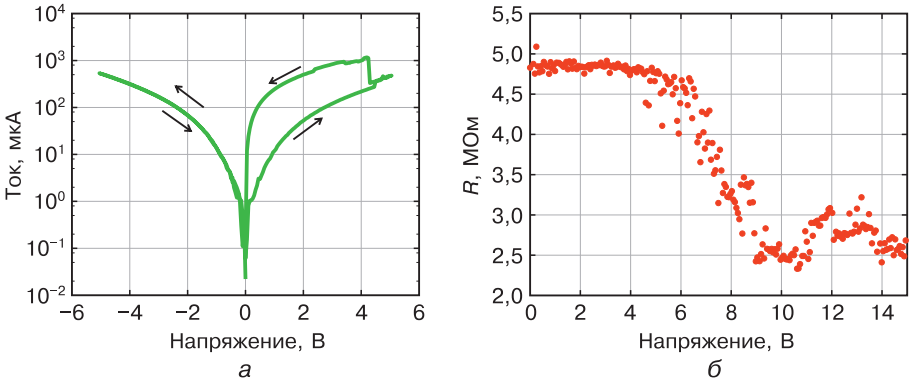



**Рис. 6.31.** Вольт-амперная характеристика (*а*) и изменение сопротивления в зависимости от амплитуды 10 мс импульса напряжения (*б*) ячейки 1D2M мемристорно-диодного кроссбара запоминающей матрицы

### 6.7.4. Численное моделирование и тестирование аппаратного импульсного персептрона

Нейросеть, состоящая из четырех виртуальных входных нейронов и двух аппаратных выходных нейронов, обучалась для выполнения задачи распознавания входных картинок с разрешением 2 × 2 пикселя. Значения яркостей пикселей преобразовывались в среднюю частоту последовательности входных импульсов с помощью микроконтроллера (рис. 6.29, *4*). Выходные импульсы активировавшихся нейронов фиксировались этим же микроконтроллером.

Моделирование производилось в программном пакете LTspice. В качестве модели мемристора использовалась модифицированная модель [15], в которой вместо постоянных сопротивлений в низкопроводящем и высокопроводящем состояниях были использованы усредненные экспериментальные ВАХ этих состояний в виде табличных функций. Скорость и пороги переключения были подстроены согласно данным на рис. 6.31, *б*. Подготовка входных данных, запуск расчетов и обработка выходных данных выполнялась Python-скриптом.

Экспериментальная и модельная кривые обучения, представленные на рис. 6.32, *а*, показывают долю правильно калссифицированных входных картинок от числа циклов (эпох) обучения. Каждый цикл обучения содержал 128 картинок, образованных путем добавления шума к исходным двум эталонным картинкам.

Модельная кривая обучения более сглаженная по сравнению с экспериментальной. Это объясняется тем, что изменение проводимости в модели мемристора происходит плавно, без скачков.

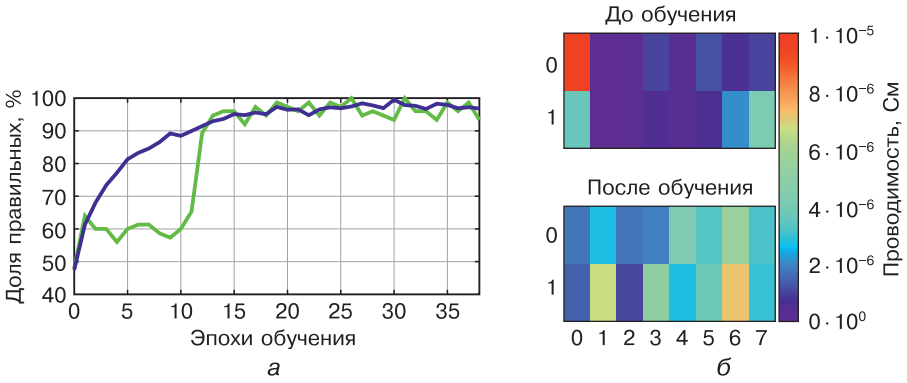



**Рис. 6.32.** Результат ассоциативного самообучения аппаратной нейросети:

*а* — изменение доли правильных классификаций: зеленая кривая — эксперимент, синия — SPICE-моделирование; *б* — изменение проводимости мемристоров кроссбара

Расположение прямоугольников на рис. 6.32, *б* соответствует положению мемристоров на рис. 6.28, *а*. Изначально проводимость мемристоров была низкой за исключением одного мемристора. В процессе ассоциативного самообучения проводимость этого мемристора уменьшилась, а остальных выросла.

При последующем переобучении нейросеть опирается на ассоциации, сформированные в процессе обработки входных сигналов, а не заложенные искусственно, как в [28]. Процесс генерации новой ассоциации при переобучении, связанным с поступлением новой информации, показан на рис. 6.33. Для демонстрации процесса быстрого формирования новой ассоциации были выбраны импульсы большой длительности, чтобы переобучение произошло за небольшое количество импульсов. Для сравнения, в [32] показан начальный и конечный результат формирования новой ассоциации после большого числа импульсов.

Совпадение выходных импульсов во времени обусловлено заданием соответствующих одинаковых входных импульсов напряжения в эксперименте и при моделировании. Различие в скорости нарастания амплитуды напряжения новых импульсов связано с тем, что изменение проводимости мемристора в эксперименте имеет вероятностный разброс (см. рис. 6.31, *б*).

Нарастание входного напряжения нейрона, которое пропорционально синаптическому току, вызвано усилением синапса при формировании ассоциации. Генерация новой ассоциации при переобучении происходит в изготовленном мемристорно-диодном кроссбаре в отличие от существующих нейросетей с синапсами на базе дискретных мемристоров [32–36].

На рис. 6.34 показана генерация новой ассоциации при подключении второго аппаратного нейрона в схему стенда.

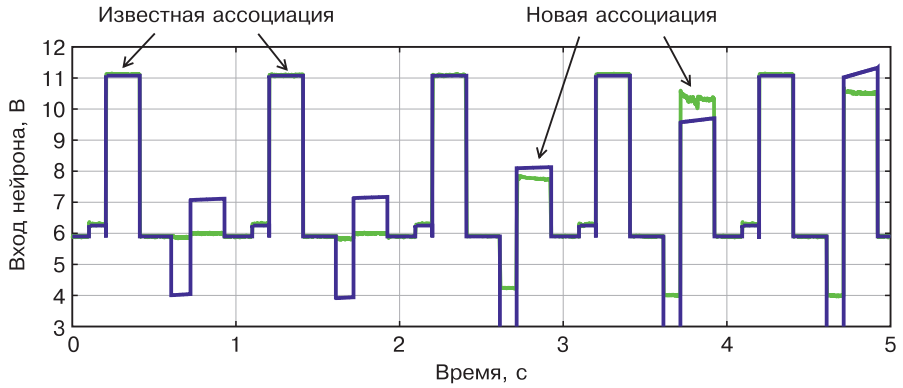



**Рис. 6.33.** Генерация новой ассоциации на фоне известной:
зеленая кривая — эксперимент; синяя — SPICE-моделирование

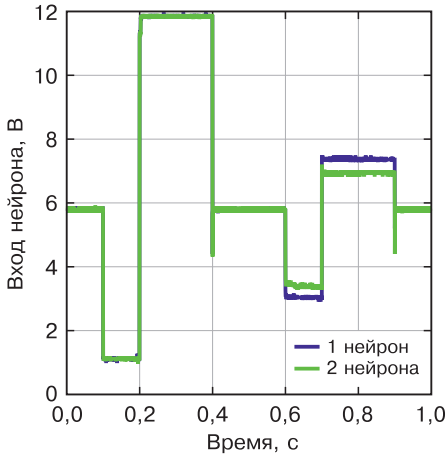



**Рис. 6.34.** Изменение амплитуды новой ассоциации
при подключении второго аппаратного нейрона в схему стенда

Подключение второго аппаратного нейрона вызывает изменение синаптического тока первого аппаратного нейрона из-за существования небольших паразитных токов через соседние ячейки в кроссбаре. Часть суммарного тока синапсов первого нейрона втекает во второй нейрон, поэтому наблюдаемая амплитуда импульсов уменьшается.

## 6.8. ВЫВОДЫ

Разработан и собран универсальный измерительный стенд, позволяющий исследовать работу мемристорно-диодных кроссбаров в качестве массива синапсов аппаратной импульсной нейросети.

Изготовлен лабораторный комбинированный мемристорно-диодный кроссбар — основа аппаратной реализации нейропроцессора. Изготовлена и протестирована импульсная нейросеть с мемристорными синапсами на основе мемристорно-диодного кроссбара и аппаратных нейронов, представляющая собой однослойный персептрон. Персептрон может рассматриваться в качестве первого слоя 3D-нейросети [31], выполняющего первичную обработку поступающей информации в биоморфном нейропроцессоре [16].

Получены экспериментальная и модельная кривые обучения, показывающие ожидаемое увеличение доли правильных классификаций с ростом числа эпох обучения. Показано влияние соседних аппаратных нейронов на синаптический ток, возникающее вследствие паразитных токов между соседними ячейками в мемристорно-диодном кроссбаре. Это влияние необходимо учитывать при разработке аппаратных нейросетей с большими кроссбарами.

Сравнение результатов моделирования и эксперимента по обучению небольшой нейросети с малым кроссбаром позволяет создать адекватные модели аппаратных нейросетей с мемристорно-диодным кроссбаром большой размерности.

Впервые продемонстрирована генерация новой ассоциации (нового знания) при обработке импульсов в изготовленном мемристорно-диодном кроссбаре в отличие от ассоциативного самообучения в существующих аппаратных нейросетях с синапсами на базе дискретных мемристоров [32–36].

Полученные экспериментальные данные свидетельствуют об эффективной работоспособности нового компонента наноэлектроники — комбинированного мемристорно-диодного кроссбара, предназначенного для изготовления запоминающей и логической матриц, входного и выходного устройств биоморфного нейропроцессора [16].

Созданы предпосылки для изготовления прототипа нейропроцессора нового поколения, качественно отличающегося от существующих нейропроцессоров на простых нейронах, предназначенных для работы компьютерного зрения, машинного обучения и других систем со слабым искусственным интеллектом.

С помощью изготовленной и протестированной аппаратной импульсной нейросети поступление новой неизвестной информации можно отождествлять с генерацией новых ассоциаций в биоморфном нейропроцессоре и при совершенствовании нейросети [31] научиться осмысливать эту информацию и, следовательно, совершить переход от слабого к сильному искусственному интеллекту.

## Список литературы

1. *Bobylev A.N.*, *Udovichenko S.Yu.* The electrical properties of memristor devices $TiN/Ti_xAl_{1-x}O_y/TiN$ produced by magnetron sputtering // Russian Microelectronics. 2016. Vol. 45. No. 6. Pp. 396–401.

2. *Писарев А.Д.* Spice-моделирование процессов ассоциативного самообучения и безусловного разобучения в логическом блоке нейропроцессора // Вестник ТюмГУ. Физико-математическое моделирование. Нефть, газ, энергетика. 2018. Т. 4. № 3. С. 132–145.
3. *Gao L.*, *Hoskins B.*, *Strukov D.* Correlation between diode polarization and resistive switching polarity in Pt/TiO2/Pt memristive device // Physca Status Solidi. RRL. 2016. Vol. 10. No. 5. Pp. 1–5.
4. *Hadiyawarman*, *Budiman F.*, *Hernowo D.G.O.* et al. Recent progress on fabrication of memristor and transistor-based neuromorphic devices for high signal processing speed with low power consumption // Japanese Journal of Applied Physics. 2018. Vol. 52. No. 3S2. P. 03EA06.
5. *Matveyev Y.*, *Kirtaev R.*, *Fetisova A.* et al. Crossbar nanoscale $HfO_2$-based electronic synapses // Nanoscale Research Letters. 2016. Vol. 11. P. 147.
6. *Klimin V.S.*, *Tominov R.V.*, *Avilov V.I.* et al. Nanoscale profiling and memristor effect of ZnO thin films for RRAM and neuromorphic devices application // Proc. SPIE 11022, International Conference on Micro- and Nano-Electronics. 2019. P. 110220E.
7. *Zhang H.*, *Gao B.*, *Sun B.* et al. Ionic doping effect in $ZrO_2$ resistive switching memory // Applied Physics Letters. 2010. Vol. 96. P. 123502.
8. *Orlov O.M.*, *Chuprik A.A.*, *Baturin A.S.* et al. Nonvolatile memory cells based on the effect of resistive switching in depth-graded ternary $Hf_xAl_{1-x}O_y$ oxide films // Russian Microelectronics. 2014. Vol. 43. No. 4. Pp. 239–245.
9. *Alekhin A.P.*, *Chouprik A.A.*, *Gudkova S.A.*, *Markeev A.M.* Structural and electrical properties of $Ti_xAl_{1-x}O_y$ thin films grown by atomic layer deposition // Journal of Vacuum Science and Technology B. 2011. Vol. 29. P. 01A302.
10. *Peng C.-S.*, *Chang W.-Y.*, *Lee Y.-H.* et al. Improvement of resistive switching stability of $HfO_2$ films with al doping by atomic layer deposition // Electrochemical and Solid-State Letters. 2012. Vol. 15. No. 4. Pp. H88–H90.
11. *Бобылев А.Н.*, *Удовиченко С.Ю.*, *Бусыгин А.Н.*, *Ибрагим А.Х.А.* Увеличение диапазона резистивного переключения мемристора для реализации большего числа синаптических состояний в нейропроцессоре // Вестник Тюменского гос. ун-та. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. Т. 5. № 2. С. 124–136.
12. *Bobylev A.N.*, *Udovichenko S.Y.*, *Busygin A.N.*, *Ebrahim A.H.* The effect of aluminum dopant amount in titania film on the memristor electrical properties // Nano Hybrids and Composites. 2020. Vol. 28. Pp. 59–64.
13. *Гудкова С.А.* Исследование структуры и свойств двух и трехкомпонентных оксидов $Ti_xAl_{1-x}O_y$, сформированных методом атомарно-слоевого осаждения: Автореф. дис. ... канд. физ.-мат. наук. МФТИ, Долгопрудный, 2011. 20 с.
14. *Алехин А.П.*, *Батурин А.С.*, *Григал И.П.* и др. Мемристор на основе смешанного оксида металлов // 2013. Патент РФ № 2472254. Патентообладатель МФТИ.
15. *Maevsky O.V.*, *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.* Complementary memristor-diode cell for a memory matrix in neuromorphic processor // International journal of nanotechnology. 2018. Vol. 15. No. 4/5. Pp. 388–393.
16. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Y.*, *Maevsky O.V.* The biomorphic neuroprocessor based on the composite memristor-diode crossbar // Microelectronics Journal. 2020. Vol. 102. P. 104827.
17. *Vinet M.*, *Batude P.*, *Tabone C.* et al. 3D monolithic integration: Technological challenges and electrical results // Microelectronic Engineering. 2011. Vol. 88. No. 4. Pp. 331–335.

18. *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. Vol. 547. Pp. 74–78.
19. *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. Vol. 176. No. 16. Pp. 1277–1284.
20. *Abe H.*, *Fujishima M.*, *Komiyama T.* et al. Heterojunction characteristics of ZnO and CuO substrates formed by direct bonding // Physica Status Solidi. C. 2012. Vol. 9. No. 6. Pp. 1396–1399.
21. *Stephen J.*, *Grimshaw J.A.* The electrical behaviour of abrupt ion implanted and diffused $p^+n$ junctions // Radiation Effects. 1971. Vol. 7. Pp. 73–85.
22. *Rubin L.*, *Poate J.* Ion implantation in silicon technology // Industrial Physicist. 2003. Vol. 9. No. 3. Pp. 12–15.
23. *Lee M.-J.*, *Park Y.*, *Kang B.-S.* et al. 2-stack ID-IR Cross-point structure with oxide diodes as switch elements for high density resistance RAM applications // IEEE International Electron Devices Meeting. 2007. Pp. 771–774.
24. *Писарев А.Д.*, *Бусыгин А.Н.*, *Бобылев А.Н.* и др. Выбор материалов и нанотехнология изготовления комбинированного мемристорного-диодного кроссбара — основы аппаратной реализации нейропроцессора // Вестник Тюменского гос. ун-та. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. Т. 5. № 4. С. 200–219.
25. *Kasap S.O.* Principles of electronic materials and devices. 4th ed. New York, NY, USA: McGraw-Hill. 2018, 978 p.
26. *Писарев А.Д.*, *Бусыгин А.Н.*, *Бобылев А.Н.* и др. Исследование электрофизических свойств комбинированного мемристорно-диодного кроссбара, являющегося основой для аппаратной реализации биоморфного нейропроцессора // Вестник Тюменского гос. ун-та. Физико-математическое моделирование. Нефть, газ, энергетика. 2020. Т. 6. № 3. С. 93–109.
27. *Gao L.*, *Hoskins B.*, *Strukov D.* Correlation between diode polarization and resistive switching polarity in Pt/$TiO_2$ /Pt memristive device // Physca Status Solidi. RRL. 2016. Vol. 10. No. 5. Pp. 1–5.
28. *Pisarev A.*, *Busygin A.*, *Bobylev A.* et al. 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. Vol. 236. P. 111471.
29. *Pisarev A.D.*, *Busygin A.N.*, *Udovichenko S.Yu.*, *Maevsky O.V.* 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor // Microelectronic Engineering.2018. Vol. 198. Pp. 1–7.
30. *Бусыгин А.Н.*, *Бобылев А.Н.*, *Губин А.А.* и др. Численное моделирование и экспериментальное исследование аппаратной импульсной нейросети с мемристорными синапсами // Вестник Тюменского гос. ун-та. Физико-математическое моделирование. Нефть, газ, энергетика. 2021. Т. 7. № 2. С. 223–235.
31. *Filippov V. A.*, *Bobylev A.N.*, *Busygin A.N.* et al. A biomorphic neuron model and principles of designing a neural network with memristor synapses for a biomorphic neuroprocessor // Neural Computing and Applications. 2020. Vol. 32. Pp. 2471–2485.
32. *Minnekhanov A.A.*, *Emelyanov A.V.*, *Lapkin D.A.* et al. Parylene based memristive devices with multilevel resistive switching for neuromorphic applications // Scientific Reports. 2019. Vol. 9. P. 10800.

33. *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. 2018. Vol. 65. No. 7. Pp. 2210–2220.
34. *Yang L.*, *Zeng Z.*, *Huang Y.*, *Wen S.* Memristor-based circuit implementations of recognition network and recall network with forgetting stages // IEEE Transactions on Cognitive and Developmental Systems. 2018. Vol. 10. No. 4. Pp. 1133.
35. *Wang Z.*, *Rao M.*, *Han J.-W.* et al. Capacitive neural network with neuro-transistors // Nature Communications. 2018. Vol. 9. P. 3208.
36. *Zhang X.*, *Long K.* Improved learning experience memristor model and application as neural network synapse // IEEE Access. 2019. Vol. 7. Pp. 15262–15271.

**Писарев Александр Дмитриевич**
**Удовиченко Сергей Юрьевич**

# Биоморфный нейропроцессор на основе наноразмерного комбинированного мемристорно-диодного кроссбара