

Александр Дмитриевич ПИСАРЕВ¹

УДК 621.382; 004.33

ЭНЕРГОЭФФЕКТИВНОЕ БИОМОРФНОЕ ИМПУЛЬСНОЕ КОДИРОВАНИЕ ИНФОРМАЦИИ В ЭЛЕКТРОННЫХ НЕЙРОНАХ ДЛЯ ВХОДНОГО БЛОКА НЕЙРОПРОЦЕССОРА

¹ кандидат технических наук, доцент
кафедры прикладной и технической физики,
заведующий лабораторией пучково-плазменных технологий
НОЦ «Нанотехнологии», Тюменский государственный университет
pscb.doc@gmail.com; ORCID: 0000-0002-5602-3880

Аннотация

Статья посвящена реализации во входном блоке нейропроцессора некоторых известных принципов информационной работы биологических систем, в том числе спайкового кодирования информации, применяемого в моделях нейронных сетей последнего поколения. Развитие современных нейросетевых ИТ порождает ряд актуальных задач, находящихся на стыке нескольких научных дисциплин. Одна из них заключается в создании аппаратной платформы — нейропроцессора для энергоэффективной работы нейросетей. Разработка нанотехнологии основных блоков нейропроцессора в последнее время ведется на основе комбинированных мемристорных сверхбольших логических и запоминающих матриц. Топология матриц построена по принципу максимальной интеграции программируемых связей между узлами. В настоящей статье описан способ реализации биоморфной нейронной функциональности на базе программируемых связей высокоинтегрированной 3D-логической матрицы.

Во введении статьи основное внимание сконцентрировано на проблеме достижения энергоэффективности работы аппаратных средств, применяемых для моделирования нейронных сетей. В основной части исследования проанализированы известные фак-

Цитирование: Писарев А. Д. Энергоэффективное биоморфное импульсное кодирование информации в электронных нейронах для входного блока нейропроцессора / А. Д. Писарев // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. Том 5. № 3. С. 186-212.
DOI: 10.21684/2411-7978-2019-5-3-186-212

ты принципов передачи и обработки информации в биологических системах с точки зрения реализации их во входном блоке нейропроцессора. В тексте рассмотрена схема электронного нейрона, реализуемая на базе элементов 3D-логической матрицы. Представлен импульсный способ кодирования входной информации, который наиболее реалистично отражает принцип работы сенсорной биологической нейронной системы. Проанализирована модель электронного нейрона для выбора диапазонов технологических параметров в реальной схеме 3D-логической матрицы. Показана реализация дизъюнктивных нормальных форм на примере работы логической функции во входном блоке нейропроцессора. Представлены результаты моделирования фрагментов электрических цепей с мемристорами 3D-логической матрицы в режиме программирования. Биоморфное импульсное кодирование стандартных цифровых сигналов позволяет достичь высокой степени энергоэффективности работы логических элементов нейропроцессора за счет уменьшения количества срабатываний вентилях. Энергоэффективность дает возможность преодолеть тепловое ограничение масштабируемой технологии трехмерной компоновки элементов в мемристорных кроссбарах.

Ключевые слова

Нанoeлектроника, нейронные сети, распределенные вычисления, биоморфный нейропроцессор, комбинированный кроссбар, мемристор.

DOI: 10.21684/2411-7978-2019-5-3-186-212

Введение

Важным научным направлением современных ИТ является интенсивно развивающаяся в последнее время область разработки программно-аппаратных средств, предназначенных для реализации функций искусственных и биоморфных нейронных сетей. Новые достижения нейросетевых ИТ весьма востребованы в хозяйственной деятельности человека, где необходима высокоскоростная обработка больших потоков информационных данных с минимальными экономическими затратами. Среди эффективно внедряемых нейросетевых ИТ можно привести в качестве примеров: работу информационных роботов в call-центрах [30], автоматизированную обработку анкетной информации в банках [12], распознавание изображений с целью выявления нарушений на производствах [1] и многие другие применения [6, 7]. Замена рутинного труда человекоподобными программно-аппаратными системами уже предоставляет ощутимый экономический эффект. В соответствии со стратегическими планами дорожной карты программы Нейронет (англ. NeuroNet, NeuroWeb, Brainet) [5] уже в ближайшее время в нейросетевых ИТ будут достигнуты новые возможности в обработке информации, в том числе связанные с получением новых знаний и реализацией сильного искусственного интеллекта.

По увеличивающемуся за последнее время количеству публикаций нейронной тематики [1, 5-7, 12, 30] можно увидеть, что к задачам нейросетевых ИТ подключаются специалисты разных научных направлений. К проблемам в этой

области всесторонне подходят нейробиологи, психологи-социологи, прикладные математики, физики, электротехники и нанотехнологи. Однако, несмотря на успехи применений нейросетей, специалистам требуется решить еще множество сложных задач, находящихся на стыке нескольких научных дисциплин. Одной из них является задача, заключающаяся в создании специализированного электронного средства, которое будет выступать в качестве аппаратной платформы для энергоэффективной и быстродействующей реализации моделей информационных и биоморфных нейросетей.

Современные информационные нейронные сети в своей основе представляют собой математические модели [19], сложность и эффективность которых возрастает от того, насколько точно закладываются в них принципы информационной работы биологической нейронной системы. Обычно для технической реализации нейросетей разработчиками выбирается аппаратная платформа, основанная на классических компьютерных системах с архитектурой фон Неймана [18]. Этот выбор в большей степени обусловлен отсутствием более подходящей аппаратной альтернативы. Разработчикам приходится адаптировать программы нейросетевых моделей, отличающихся сильно распределенными вычислениями, к архитектуре фон Неймана с последовательным циклом выполнения команд. Ситуация осложняется тем, что искусственные модели нейросетей требуют свойства сверхмасштабируемости. Выполнение этих условий является сложнейшей алгоритмической задачей с применением последних достижений в объектно-ориентированном программировании. Такой подход нецелесообразен, поскольку концепция применения нейросетей должна упрощать программирование передовых систем, заменяя квалифицированный труд написания программ на более простое глубокое обучение искусственных нейросетей.

Оптимизации фон-неймановских аппаратных систем для нейросетевых применений посвящена работа [18], из которой следует, что на базе существующих аппаратных универсальных платформ, в том числе с применением современных ускорителей векторно-матричных операций, разработчикам не достичь качественного прорыва в скорости и энергоэффективности работы нейросетевых программ. Существует проблема, которая принципиально ограничивает эффективность расчета сверхбольших нейросетей на фон-неймановских аппаратных архитектурах. Этот непреодолимый недостаток аппаратных средств известен как «узкое горлышко» архитектуры фон Неймана [4]. Суть проблемы заключается в том, что огромное количество одновременно вычисляемых функций сверхбольшой совокупности нейронов (обычно требуется до 10^9 нейронов в случае эмуляции функций человеческой центральной нервной системы) приводит к увеличению времени работы в нелинейно возрастающей прогрессии. Как следствие, возникает быстрое насыщение вычислительных возможностей аппаратной системы при стремительно нарастающем энергопотреблении.

Классическая системная плата с многоядерной процессорной и программируемой логической архитектурами выполняет множество энергозатратных операций с движением (копированием) данных в памяти между своими блока-

ми. Для сверхбольших нейросетей требуется множество массивных охлаждаемых процессорных модулей и дорогие суперкомпьютеры. Моделирование работы огромных нейросетей с высокой точностью могут себе позволить только гиганты индустрии ИТ (Google, Microsoft). Повышение точности работы многих других нейросетевых приложений оказывается экономически неоправданно по причине низкой энергоэффективности вычислительного оборудования и значительных аппаратных затрат.

Становится все более очевидно, что неподходящее базовое оборудование, которое сейчас основано на классических компьютерах, ограничивает возможности и дальнейшее развитие применений нейросетевых технологий. Возрастающая вычислительная сложность нейронных моделей требует повышения быстродействия и энергоэффективности вычислительных платформ, что невозможно сделать без качественного скачка в оптимизации аппаратных ресурсов. Таким образом, разработка специализированного энергоэффективного и быстродействующего аппаратного средства — нейропроцессора является ключевой научно-технической задачей нейросетевых ИТ.

Разработка нейропроцессора на основе мемристорного кроссбара начата в работе [16]. Полная концепция нейропроцессора для реализации биоморфной нейросети предложена в публикации [13]. Основная идея разработки заключается в применении унифицированной логической [32] и запоминающей [8] матриц, состоящих из комбинированных КМОП-мемристорных элементов, в блоках нейропроцессора. Представленные матрицы разработаны с учетом требования высокой интеграции программируемых коммутационных связей между элементами логических сумматоров. Топология и нанотехнология изготовления запоминающей и логической матриц описаны в статье [14]. Слои сумматоров в матрицах объединены в соответствии с биоморфным принципом коммутации для реализации связей нейронов слоями по принципу «каждый с каждым». Такая топология позволяет организовать процессы самообучения и разобучения в нейропроцессоре, представленные в работе [11].

Входной блок нейропроцессора предназначен для согласования и ввода информации в главные его блоки. Этот модуль требует первостепенной разработки, поскольку в нем закладывается первичный способ кодирования информации, который будет использоваться в остальных блоках нейропроцессора. Работа всего нейропроцессора основана на принципах работы биологических нейронных систем. Есть некоторые отличия работы входного блока от других блоков нейропроцессора. Они заключаются в том, что во входном блоке не требуется реализация механизмов памяти, поэтому многие сложные биологические информационные процессы можно не учитывать. Для первичного преобразования во входном блоке нейропроцессора необходимо выполнять следующие функции: кодирование информации, сжатие и фильтрацию данных. Принципы сжатия и фильтрации данных для нейропроцессора описаны в работе [10]. Настоящая статья посвящена биоморфному кодированию информации.

Принципы импульсного кодирования информации в биологических системах

В основу импульсного кодирования во входном блоке нейропроцессора с целью достижения высокой энергоэффективности задумано использовать известные принципы информационной работы биологических сенсорных нейронов. Стоит отметить, что многие аспекты работы биологической сенсорной системы еще не получили точного научного объяснения. Однако на сегодняшний день достоверно установлены некоторые принципы, являющиеся достаточно простыми и информационно-функциональными. Эти принципы будут рассмотрены 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, называемой в нейробиологии потенциалом действия, который был впервые исследован в работах А. Ходжкина и Э. Хаксли [24]. 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 (англ. spike) [34]. 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 (англ. spike-timing-dependent plasticity, или в переводе — пластичность, зависящая от времени спайка), описанному в работе [34]. Однако во входном блоке 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 следующему 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* [20]. *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 работами, изучающими скорость передачи тактильных событий с кончика пальца человека во время манипуляций с объектом [25], или исследованиями по прохождению информационных сигналов от нейронов сетчатки глаза при кратко проецируемых на нее изображениях [21]. 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 [28, 33, 35].

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 м/с до 100 м/с в зависимости от размера волокна и наличия миелинизированной оболочки [2]). 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 [36] 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 [20, 27]. 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 (англ. *spiking neural network*, SNN) [22, 26]. SNN являются биоморфными нейронными сетями третье-

го поколения, которые наиболее реалистично на сегодняшний день отражают работу биологической нейронной системы. Предложенные правила SPAN [29] и PSD [37] 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 нейропроцессора, проходит через несколько слоев входных нейронов, где осуществляется кодирование, сжатие и фильтрация 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 блоке 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 указанные биоморфные принципы могут способствовать высокой энергоэффективности нейропроцессора в случае воплощения их во входном блоке.

Основа электронного нейрона

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. Как было показано в работе [10], импульсное кодирование данных во входном блоке нейропроцессора является эффективным для операций сжатия и фильтрации информационных данных.

На рис. 1а представлена принципиальная схема электронного нейрона. Электронный нейрон построен с использованием минимального количества элементов. Так сделано для того, чтобы простые электронные схемы, имитирующие нейроны, можно было объединить в сверхбольшие массивы и реализовать в объеме 3D-логической матрицы, топология и технология которой разработана ранее в [14].

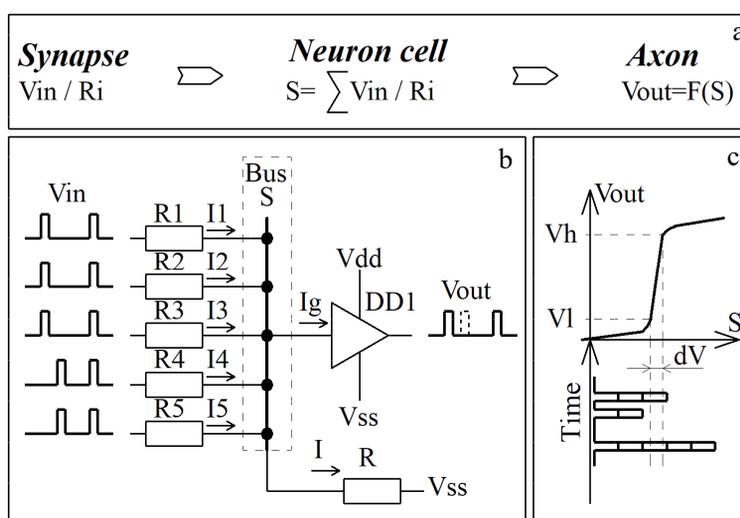


Рис. 1. Принцип суммации импульсов в электронном нейроне:
 а) функциональная схема информационного нейрона;
 б) принципиальная схема электронного нейрона;
 в) передаточная характеристика повторителя DD1, выполняющего пороговую функцию

Fig. 1. The principle of summation of pulses in an electronic neuron:
 а) a functional diagram of an information neuron;
 б) principle scheme of an electronic neuron;
 в) transfer characteristic of the repeater DD1, performing a threshold function

На рис. 1а показаны блоки классического информационного нейрона, взвешивающего, суммирующего и сравнивающего результат с порогом. На рис. 1б показаны элементы схемы предлагаемого электронного нейрона, имитирующего блоки классического нейрона. Резисторы R1-R5 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 DD1, который имитирует выход аксона с пороговой функцией активации. На рис. 1с показана передаточная характеристика повторителя, которая имеет ступенчатую форму в области порога срабатывания в середине диапазона питания, что аналогично функциональности блока аксона в информационном нейроне.

В качестве примера входной информации, подаваемой на электронный нейрон, представлены два некоторых 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 (рис. 1б) входного напряжения V_{in} . 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 dV (рис. 1с) на суммирующей шине происходит переключение выхода инвертора DD1 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.

Условия формирования биоморфных импульсов на шинах 3D-логической матрицы

В отличие от упрощенной резистивной схемы (рис. 1), 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. Практические мемристоры как переключатели не отличаются большим изменением сопротивления (скачок сопротивления может быть в диапазоне 10^3 - 10^6 Ом). Их работа характеризуется гистерезисом вольт-амперной характеристики мемристорного материала. Ряд исследований по материаловедению направлен на достижение высокого скачка сопротивления мемристоров [3, 17, 23, 31] и получение несколь-

ких устойчивых состояний проводимости, которые могут играть роль аналога биологических синаптических состояний, как показано в работе [17]. Нанотехнология увеличения синаптических состояний мемристорных связей путем подбора новых материалов на базе смешанных оксидов представлена в статье [3]. Стоит задача физического моделирования, заключающаяся в получении максимально возможного числа связей, имитирующих синапсы, с учетом утечек и «неидеальности» параметров элементов в реальной микросхеме.

На рис. 2 показана схема фрагмента 3D-логической матрицы, полная ее конструкция представлена в работе [8]. Топология 3D-логической матрицы разработана с учетом максимальной интеграции мемристоров, приходящихся на КМОП-транзисторную инверторную структуру. В такой топологии на одной шине потенциально могут участвовать в суммировании около 100-1 000 сигналов.

Импульсные сигналы поступают на общую шину V_g через входные инверторы и мемристорные ячейки. На схеме рис. 2 мемристорные ячейки объединены в группы по общему режиму работы. Выделены четыре 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.

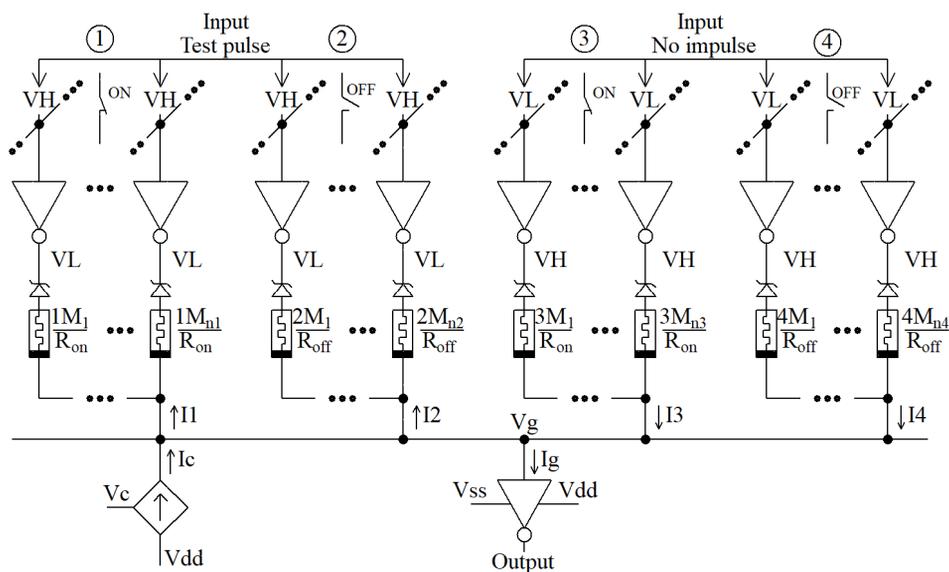


Рис. 2. Фрагмент электрической схемы 3D-логической матрицы

Fig. 2. Fragment of an electrical circuit 3D logical matrix

- 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.

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

$$-\sum_{i=1}^{n1} \left(\frac{Vg-VL_i}{R_{1M_i}^{on}+R_z^+} \right) - \sum_{i=1}^{n2} \left(\frac{Vg-VL_i}{R_{2M_i}^{off}+R_z^+} \right) + \sum_{i=1}^{n3} \left(\frac{VH_i-Vg}{R_{3M_i}^{on}+R_z^-} \right) + \sum_{i=1}^{n4} \left(\frac{VH_i-Vg}{R_{4M_i}^{off}+R_z^-} \right) + Ic = 0, \quad (1)$$

где Vg — напряжение на шине, относительно точки нижнего потенциала Vss; VH_i и VL_i — уровни напряжения, соответствующие логическим единице и нулю, которые формируют КМОП-инверторы; Ic — поступающий ток в шину от источника, установленного на периферии матрицы и предназначенного для управления порогом срабатывания КМОП-инвертора; R_{1M_i}^{on}, R_{2M_i}^{off}, R_{3M_i}^{on} и R_{4M_i}^{off} — резистивные состояния мемристоров, разделенные по четырем группам; R_z⁺, R_z⁻ — сопротивления диодов Зенера в прямом и обратном направлениях прохождения тока; n1, n2, n3 и n4 — полное количество мемристоров в группах с одинаковым режимом работы.

В сумме (1) первые четыре компонента соответствуют группам мемристоров, работающих в одинаковых условиях, показанных на схеме (рис. 2). 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 (1) можно снять суммирующие операторы с учетом введенных обозначений, определяемых параметрами 3D-логической матрицы: R_{1M_i}^{on}+R_z⁺= $\frac{1}{\sigma_{on}^+}$; R_{2M_i}^{off}+R_z⁺= $\frac{1}{\sigma_{off}^+}$; R_{3M_i}^{on}+R_z⁻= $\frac{1}{\sigma_{on}^-}$; R_{4M_i}^{off}+R_z⁻= $\frac{1}{\sigma_{off}^-}$; VH_i = VH; VL_i = VL. Через введенные параметры формулу (1) можно записать в виде:

$$-n1 \cdot \sigma_{on}^+ \cdot (Vg-VL) - n2 \cdot \sigma_{off}^+ \cdot (Vg-VL) + n3 \cdot \sigma_{on}^- \cdot (VH-Vg) + n4 \cdot \sigma_{off}^- \cdot (VH-Vg) + Ic = 0.$$

Из последнего равенства можно выразить электрическое напряжение на шине:

$$Vg = \frac{VL(n1 \cdot \sigma_{on}^+ + n2 \cdot \sigma_{off}^+) + VH(n3 \cdot \sigma_{on}^- + n4 \cdot \sigma_{off}^-) + Ic}{\sigma_{full}}, \quad (2)$$

где $\sigma_{full} = n1 \cdot \sigma_{on}^+ + n2 \cdot \sigma_{off}^+ + n3 \cdot \sigma_{on}^- + n4 \cdot \sigma_{off}^-$ — 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 $k \cdot (VH - VL)$, где k — безразмерный коэффициент, определяемый уровнем напряжения переключения по передаточной характеристике инвертора и имеющий обычно значение около 0,5.

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

$$k \cdot 0,5(VH - VL) > \frac{VL(n1 \cdot \sigma_{on}^+ + n2 \cdot \sigma_{off}^+) + VH(n3 \cdot \sigma_{on}^- + n4 \cdot \sigma_{off}^-) + Ic}{\sigma_{full}}. \quad (3)$$

Представленное условие (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.

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 (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 (3) можно легко адаптировать для мемристоров, имеющих несколько устойчивых состояний электрической проводимости.

Источник тока Ic для управления порогом срабатывания можно заменить на резистор с проводимостью $\sigma_c = Ic / (VH - Vg)$. В этом случае, предполагая, что $VL = 0$, формулу (4) можно переписать в виде:

$$Vg(n1, n2) = \frac{VH(n3 \cdot \sigma_{on}^- + n4 \cdot \sigma_{off}^- + \sigma_c)}{n1 \cdot \sigma_{on}^+ + n2 \cdot \sigma_{off}^+ + n3 \cdot \sigma_{on}^- + n4 \cdot \sigma_{off}^- + \sigma_c}. \quad (4)$$

Зависимость (4) для напряжения на шине представлена на графиках (рис. 3), построенных при двух значениях проводимости σ_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 2 000 мВ. Изначально предполагается, что в шину входят 10^5 мемристорных связей всех режимов.

Функция $V_g(n1, n2)$ на рис. 3 показана семейством кривых, полученных путем фиксирования одного из параметров. Мемристорные связи, работающие в режиме с обратносмещенным диодом Зенера (их количество обозначено $n3$ и $n4$), не оказывают на кривые графика заметного влияния из-за высокого электрического сопротивления. Сопротивление диода Зенера составляет 10^2 Ом и 10^9 Ом в прямом и обратном направлениях соответственно. Поэтому увеличение количества двух оставшихся типов связей $n1$ и $n2$ производилось за счет пренебрежимо малого изменения $n3$ и $n4$.

Можно видеть на графиках рис. 3, 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

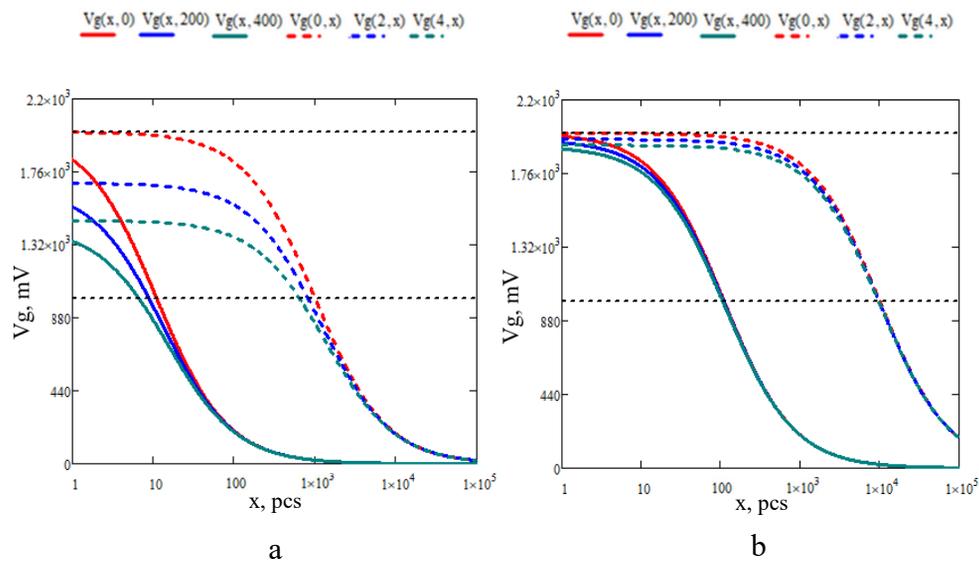


Рис. 3. Зависимости напряжения на шине от количества мемристоров первой и второй группы во время подачи входных импульсов:

- а) графики построены для $\sigma_c = 10^{-2}$ 1/Ом;
 б) графики построены для $\sigma_c = 10^{-1}$ 1/Ом

Fig. 3. Dependences of the bus voltage on the number of memristors of the first and second groups during the supply of input pulses:

- a) graphs are plotted for $\sigma_c = 10^{-2}$ 1/Ом;
 b) graphs plotted for $\sigma_c = 10^{-1}$ 1/Ом

срабатывания инвертора задан на уровне $0,5 \cdot V_H$. 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 σ_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. 3. Ложный выходной импульс может появляться в результате суммирования чрезмерного количества входных импульсов, поступающих на «закрытые» мемристорные связи. Нежелательные срабатывания могут быть ограничены допустимым количеством входных импульсов. Эта величина определяется диапазоном резистивного переключения мемристорного материала.

На графике рис. 3а видно значительное расхождение кривых, соответствующих проводимости $\sigma_c = 10^{-2}$ 1/Ом, по сравнению с кривыми на рис. 3б 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 $0,5 \cdot V_H$, условие прохождения информационного импульса определяется неравенством:

$$\sigma_c(n1, n2) < n1 \cdot \sigma_{он}^+ + n2 \cdot \sigma_{от}^+ \quad (5)$$

По условию (5) на рис. 4 построены области, соответствующие режимам работы электронного нейрона в 3D-логической матрице. Двухпараметрические зависимости показаны на графике путем фиксирования одного из параметров.

График на рис. 4а 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 «отсутствие импульсов» и «информационные импульсы» (рис. 4а). Область «ложные импульсы» ограничивает количество связей для установленного порога срабатывания выходного инвертора.

График на рис. 4б 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.

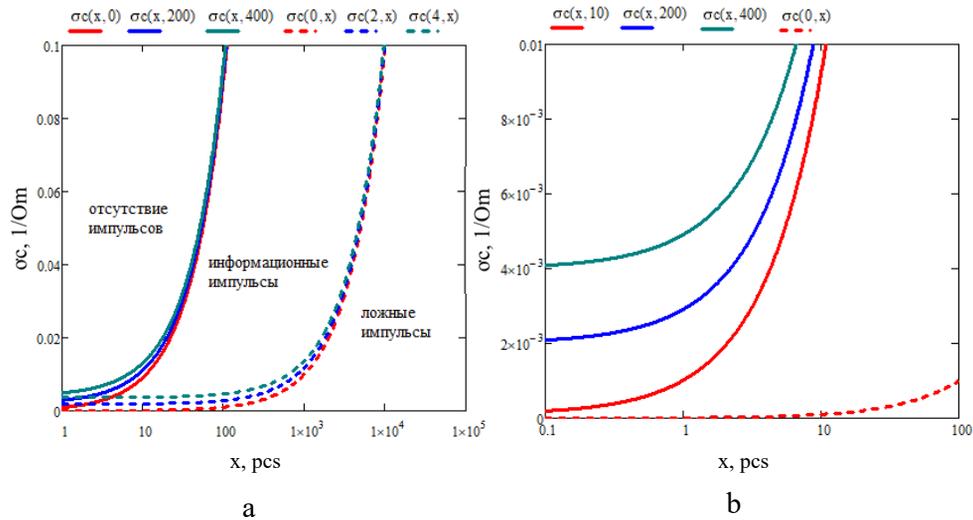


Рис. 4. Зависимости проводимости эквивалента источника тока для управления порогом срабатывания от количества мемристоров первой и второй группы в двух режимах работы электронного нейрона: а) режим суммации импульсов; б) логический режим

Fig. 4. Dependences of the conductivity equivalent of the current source for controlling the response threshold on the number of memristors of the first and second groups in two modes of operation of the electronic neuron: а) pulse summation mode; б) logical mode

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 дизъюнктивные 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» (англ. «XOR»). На рис. 5 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 логи-

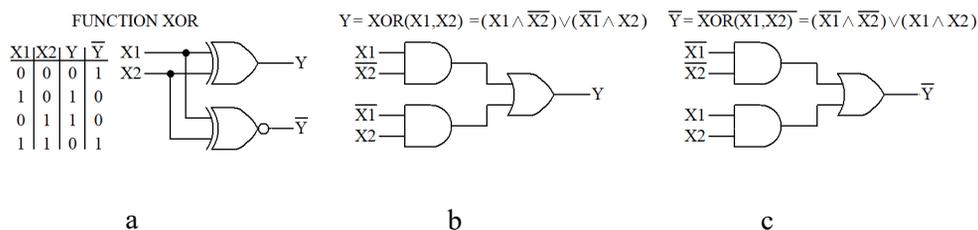


Рис. 5. Функции «XOR» с прямым и инверсным выходом:
 а) таблица истинности;
 б) СДНФ для прямого выхода;
 в) СДНФ для инверсного выхода

Fig. 5. «XOR» functions with direct and inverse output:
 а) truth table;
 б) PDNF for direct exit;
 в) PDNF for inverse output

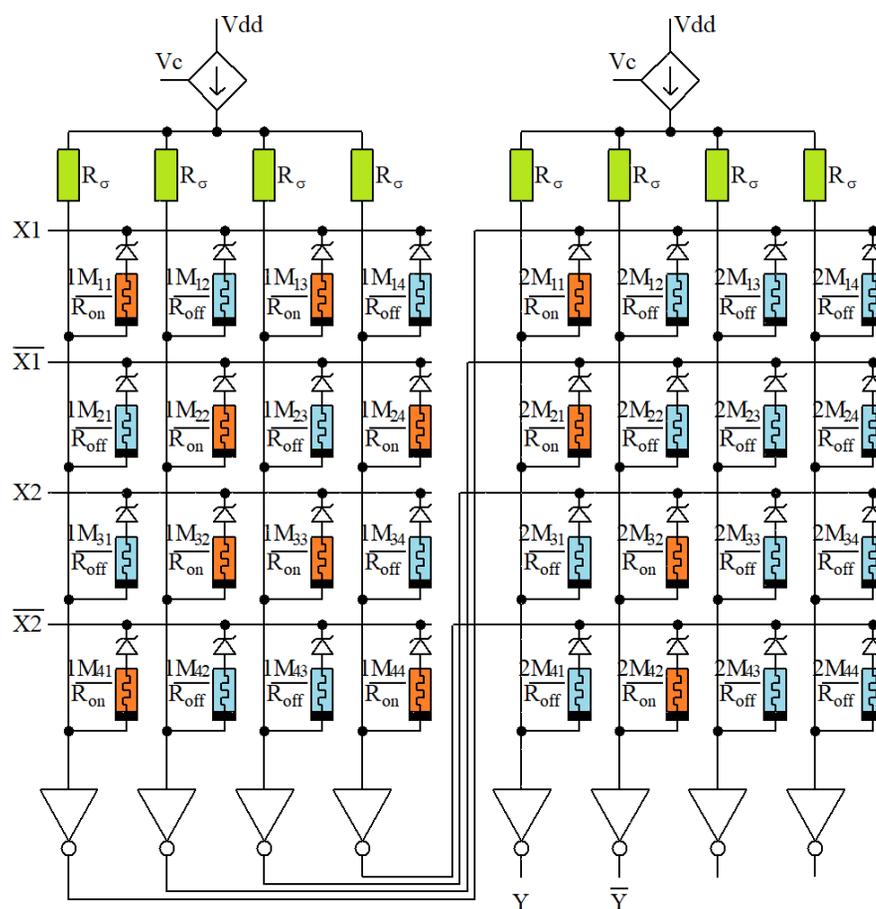


Рис. 6. Пример реализации логической функции исключающего «ИЛИ» в двух слоях 3D-логической матрицы

Fig. 6. An example of the implementation of an exclusive «OR» logical function in two layers of a 3D logical matrix

ческих схем представлены совершенные ДНФ (СДНФ). На входы схем поданы логические переменные X_1 и X_2 в прямом и инверсном видах. Как видно, 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:

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

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:

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

что эквивалентно формуле:

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

Для реализации функций «XOR» с прямым и инверсным выходами в соответствии с их СДНФ, представленной на рис. 5, мемристорные матрицы нужно определить значениями:

$$1M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad 2M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (6)$$

На рис. 6 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 красным цветами соответственно. Можно отметить, что «включенные» и «выключенные» мемристорские связи соответствуют расположению нулей и единиц в значениях $1M$ и $2M$ в равенстве (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.

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

Мемристоры 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 $R2$ показано на рис. 7. Для программирования используется инвертор, подключенный своим выходом к шине кроссбара (как показано на схеме рис. 7а), который подачей высокого уровня напряжения на управляющий вход op/wt переведен в рабочий режим из высокоомного z -состояния. В начале диаграммы (см. рис. 7b) с помощью импульса ERASE на эдюре сигнала DL выполнено изменение сопротивления мемристора из низкоомного в высокоомное состояние для «стирания» соответствующей связи. После этого на диаграмме показана подача импульса WRITE на эдюре сигнала $a2$ и переход мемристора в низкоомное состояние для создания высокопроводящей связи.

Для моделирования изменений сопротивления мемристора была использована модель $R2$, представленная в публикации [15], с параметрами $R_{on} = 10$ k; $R_{off} = 1000$ k; $R_{init} = 1000$ k; $\beta = 3e11$; $V_t = 2,2$; $C_{int} = 0,25$ p.

Диоды Зенера $D1$ и $D2$ используются в схеме 3D-логической матрицы как элемент для энергоэффективного выполнения логических функций. Принцип его работы в этом случае описан в исследовании [9]. Во время программирования состояний мемристоров диод Зенера находится в проводящем состоянии в обоих направлениях прохождения тока.

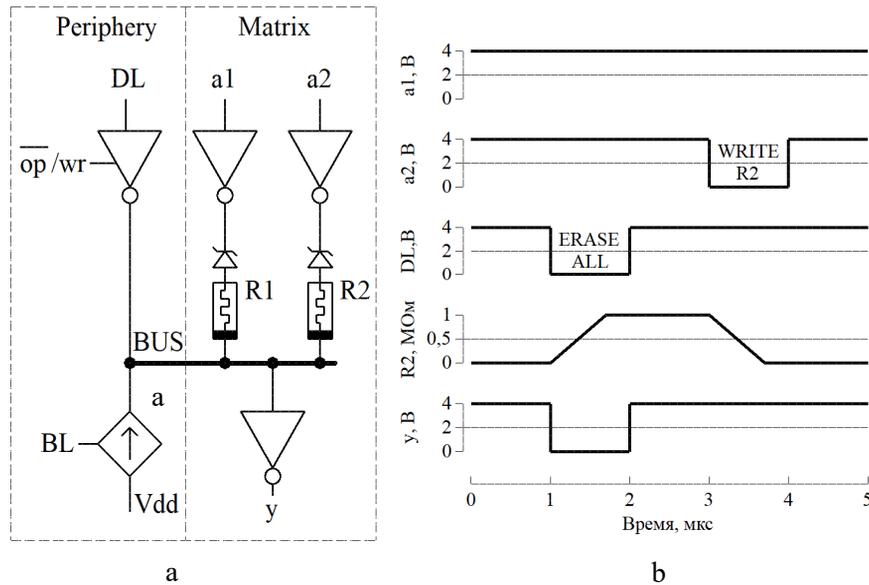


Рис. 7. Стирание и создание связи в 3D-логической матрице:
 а) схема моделирования;
 б) эмпоры сигналов и временные зависимости параметров схемы

Fig. 7. Erasing and creating a connection in a 3D logical matrix:
 а) modeling scheme;
 б) diagrams of signals and time dependences of circuit parameters

Заключение

В статье показаны принципы работы логических и биоморфных функций во входном блоке нейропроцессора, разрабатываемом на базе 3D-логической матрицы. Цифровые логические функции используются для первичного согласования стандартной информации с электронной нейронной сетью, имитирующей работу биологических сенсорных нейронов. Цифровые функции реализуются в архитектуре 3D-логической матрицы за счет полного логического базиса. Полный логический базис образуется при условии, что цифровые переменные будут подаваться на вход матрицы в прямом и инверсном виде. Для реализации любой СДНФ представлена формула, описывающая работу слоев 3D-логической матрицы, и показано моделирование процесса программирования мемристивных связей.

Суть предложенного в настоящей статье 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. В отличие от электрических импульсов спайки распространяются по нейрональным волокнам с относительно невысокой скоростью (в среднем от 1 м/с до 100 м/с в зависимости от размера волокна и наличия миелинизированной оболочки). 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 блоке 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.

Таким образом, для первичного преобразования сигналов во входном блоке нейропроцессора необходимо выполнять: кодирование информации, сжатие и фильтрацию данных. Мемристорные связи, определяющие логику работы входного блока нейропроцессора, не требуют обучения в процессе работы и программируются отдельно внешними устройствами во время создания архитектуры этого блока.

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 основное 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.

СПИСОК ЛИТЕРАТУРЫ

1. Барский А. Б. Нейронные сети: распознавание, управление, принятие решений / А. Б. Барский. М.: Финансы и статистика, 2004. 176 с.
2. Блум Ф. Мозг, разум и поведение / Ф. Блум, А. Лейзерсон, Л. Хофстедтер; пер. с англ. М.: Мир, 1988. 248 с.
3. Бобылев А. Н. Увеличение диапазона резистивного переключения мемристора для реализации большего числа синаптических состояний в нейропроцессоре / А. Н. Бобылев, С. Ю. Удовиченко, А. Н. Бусыгин, А. Х. Ибрагим // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. Том 5. № 2. С. 124-136. DOI: 10.21684/2411-7978-2019-5-2-124-136
4. Бэкус Дж. Можно ли освободить программирование от стиля фон-Неймана? Функциональный стиль и соответствующая алгебра программ / Дж. Бэкус // Лекции лауреатов премии Тьюринга. М.: Мир, 1993. 560 с.
5. Глобальное будущее 2045. Конвергентные технологии (НБИКС) и трансгуманистическая эволюция / под ред. Д. И. Дубровского. М.: Изд-во МБА, 2013. 272 с.
6. Гусев А.В. Перспективы нейронных сетей и глубокого машинного обучения в создании решений для здравоохранения / А.В. Гусев // Врач и информационные технологии. 2017. № 3. С. 92-105.
7. Каличкин В. К. Применение нейронной экспертной системы для классификации эрозионных земель / В. К. Каличкин, А. И. Павлова // Сибирский вестник с.-х. науки. 2014. № 6. С. 5-11.
8. Многослойная логическая матрица на основе мемристорной коммутационной ячейки: заявка на изобретение № 2017122704 / А. Д. Писарев, О. В. Маевский, А. Н. Бусыгин, С. Ю. Удовиченко; заявл. 27.06.2017; решение о выдаче патента 10.04.2018.
9. Писарев А. Д. Комбинированный мемристорно-диодный кроссбар как основа запоминающего устройства / А. Д. Писарев, А. Н. Бусыгин, А. Н. Бобылев, С. Ю. Удовиченко // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2017. Том 3. № 4. С. 142-149. DOI: 10.21684/2411-7978-2017-3-4-142-149
10. Писарев А. Д. Реализация дискретного косинусного преобразования во входном блоке мемристорного нейропроцессора / А. Д. Писарев // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2019. Том 5. № 1. С. 147-161. DOI: 10.21684/2411-7978-2019-5-1-147-161
11. Писарев А. Д. SPICE-моделирование процессов ассоциативного самообучения и безусловного разобучения в логическом блоке нейропроцессора / А. Д. Писарев // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2018. Том 4. № 3. С. 132-145. DOI: 10.21684/2411-7978-2018-4-3-132-145
12. Трофимова Е. А. Нейронные сети в прикладной экономике: учеб. пособие / Е. А. Трофимова, В. Д. Мазуров, Д. В. Гилёв; под общ. ред. Е. А. Трофимовой; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. Екатеринбург: Изд-во Урал. ун-та, 2017. 96 с.

13. Удовиченко С. Ю. Нейропроцессор на основе комбинированного мемристорно-диодного кроссбара / С. Ю. Удовиченко, А. Д. Писарев, А. Н. Бусыгин, О. В. Маевский // *Наноиндустрия*. 2018. № 5 (84). С. 344-355.
DOI: 10.22184/1993-8578.2018.84.5.344.355
14. Удовиченко С. Ю. 3D КМОП-мемристорная нанотехнология создания логической и запоминающей матриц нейропроцессора / С. Ю. Удовиченко, А. Д. Писарев, А. Н. Бусыгин, О. В. Маевский // *Наноиндустрия*. 2017. № 5 (76). С. 26-34.
DOI: 10.22184/1993-8578.2017.76.5.26.34
15. Biolek D. Reliable SPICE simulations of memristors, memcapacitors and meminductors / D. Biolek, M. Di Ventra, Y. V. Pershin // *Radioengineering*. 2013. Vol. 22. № 4. Pp. 945-968.
16. Bobylev A. N. Neuromorphic coprocessor prototype based on mixed metal oxide memristors / A. N. Bobylev, A. N. Busygin, A. D. Pisarev, S. Yu. Udovichenko, V. A. Filippov // *International Journal of Nanotechnology*. 2017. Vol. 14. No 7/8. Pp. 698-704. DOI: 10.1504/IJNT.2017.083444
17. Bobylev A. N. The electrical properties of memristor devices $TiN/Ti_xAl_{1-x}O_y/TiN$ produced by magnetron sputtering / A. N. Bobylev, S. Yu. Udovichenko // *Russian Microelectronics*. 2016. Vol. 45. № 6. Pp. 396-401. DOI: 10.1134/S1063739716060020
18. Chen Y.-H. Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks / Y.-H. Chen, T. Krishna, J. S. Emer, V. Sze // *IEEE Journal of Solid-State Circuits*. 2017. Vol. 52. № 1. Pp. 127-138. DOI: 10.1109/JSSC.2016.2616357
19. Filippov V. A. A biomorphic neuron model and principles of designing a neural network with memristor synapses for a biomorphic neuroprocessor / V. A. Filippov, A. N. Bobylev, A. N. Busygin, A. D. Pisarev, S. Yu. Udovichenko // *Neural Computing and Applications*. 2019. DOI: 10.1007/s00521-019-04383-7
20. Gardner B. Supervised learning in spiking neural networks for precise temporal encoding / B. Gardner, A. Grüning // *PLoS ONE*. 2016. Vol. 11. № 8. Article e0161335. Pp. 1-28. DOI: 10.1371/journal.pone.0161335
21. Gollisch T. Rapid neural coding in the retina with relative spike latencies / T. Gollisch, M. Meister // *Science*. 2008. Vol. 319. № 5866. Pp. 1108-1111.
DOI: 10.1126/science.1149639
22. Gütig R. To spike, or when to spike? / R. Gütig // *Current Opinion in Neurobiology*. 2014. Vol. 25. Pp. 134-139. DOI: 10.1016/j.conb.2014.01.004
23. Hadiyawarman. Recent progress on fabrication of memristor and transistor-based neuromorphic devices for high signal processing speed with low power consumption / Hadiyawarman, F. Budiman, D. G. O. Hernowo, R. R. Pandey, H. Tanaka // *Japanese Journal of Applied Physics*. 2018. Vol. 52. № 3S2. Article 03EA06.
DOI: 10.7567/JJAP.57.03EA06
24. Hodgkin A. L. A quantitative description of membrane current and its application to conduction and excitation in nerve / A. L. Hodgkin, A. F. Huxley // *The Journal of Physiology*. 1952. Vol. 117. № 4. Pp. 500-544. DOI: 10.1113/jphysiol.1952.sp004764
25. Johansson R. S. First spikes in ensembles of human tactile afferents code complex spatial fingertip events / R. S. Johansson, I. Birznieks // *Nature Neuroscience*. 2004. Vol. 7. № 2. Pp. 170-177. DOI: 10.1038/nn1177

26. Kasinski A. Comparison of supervised learning methods for spike time coding in spiking neural networks / A. Kasinski, F. Ponulak // *International Journal of Applied Mathematics and Computer Science*. 2006. Vol. 16. № 1. Pp. 101-113.
27. Larkum M. E. Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons / M. E. Larkum, J. J. Zhu, B. Sakmann // *The Journal of Physiology*. 2001. Vol. 533. № 2. Pp. 447-466. DOI: 10.1111/j.1469-7793.2001.0447a.x
28. Mainen Z. F. Reliability of spike timing in neocortical neurons / Z. F. Mainen, T. J. Sejnowski // *Science*. 1995. Vol. 268. № 5216. Pp. 1503-1506. DOI: 10.1126/science.7770778
29. Mohemmed A. SPAN: Spike pattern association neuron for learning spatio-temporal spike patterns / A. Mohemmed, S. Schliebs, S. Matsuda, N. Kasabov // *International Journal of Neural Systems*. 2012. Vol. 22. № 04. Article 1250012. DOI: 10.1142/S0129065712500128
30. Pavlova A. I. The use of neural networks in forecasting the load of the call-center / A. I. Pavlova, D. A. Morozov // *International Journal of Advanced Studies*. 2017. Vol. 7. № 2-2. Pp. 108-112. DOI: 10.12731/2227-930X-2017-2-2-108-112
31. Peng C.-S. Improvement of resistive switching stability of HfO₂ films with Al doping by atomic layer deposition / C.-S. Peng, W.-Y. Chang, Y.-H. Lee, M.-H. Lin, F. Chen, M.-J. Tsai // *Electrochemical and Solid-State Letters*. 2012. Vol. 15. № 4. Pp. H88-H90. DOI: 10.1149/2.011204esl
32. Pisarev A. D. 3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor / A. D. Pisarev, A. N. Busygin, S. Yu. Udovichenko, O. V. Maevsky // *Microelectronic Engineering*. 2018. Vol. 198. Pp. 1-7. DOI: 10.1016/j.mee.2018.06.008
33. Reich D. S. Response variability and timing precision of neuronal spike trains *in vivo* / D. S. Reich, J. D. Victor, B. W. Knight, T. Ozaki, E. Kaplan // *Journal of Neurophysiology*. 1997. Vol. 77. № 5. Pp. 2836-2841. DOI: 10.1152/jn.1997.77.5.2836
34. Takeuchi T. The synaptic plasticity and memory hypothesis: encoding, storage and persistence / T. Takeuchi, A. J. Duzkiewicz, R. G. M. Morris // *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2014. Vol. 369. № 1633. DOI: 10.1098/rstb.2013.0288
35. Uzzell V. J. Precision of spike trains in primate retinal ganglion cells / V. J. Uzzell, E. J. Chichilnisky // *Journal of Neurophysiology*. 2004. Vol. 92. № 2. Pp. 780-789. DOI: 10.1152/jn.01171.2003
36. Van Rullen R. Spike times make sense / R. Van Rullen, R. Guyonneau, S. J. Thorpe // *Trends in Neurosciences*. 2005. Vol. 28. № 1. Pp. 1-4. DOI: 10.1016/j.tins.2004.10.010
37. Yu Q. Precise-spike-driven synaptic plasticity: learning hetero-association of spatiotemporal spike patterns / Q. Yu, H. Tang, K. C. Tan, H. Li // *PLoS ONE*. 2013. Vol. 8. № 11. Article e78318. DOI: 10.1371/journal.pone.0078318

Alexander D. PISAREV¹

UDC 621.382; 004.33

**ENERGY EFFICIENT BIOMORPHIC PULSE
INFORMATION CODING IN ELECTRONIC NEURONS
FOR THE ENTRANCE UNIT OF THE NEUROPROCESSOR**

¹ Cand. Sci. (Tech.), Associate Professor,
Department of Applied and Technical Physics,
Head of Laboratory of Beam-Plasma Technologies,
SEC “Nanotechnologies”, University of Tyumen
spcb.doc@gmail.com; ORCID: 0000-0002-5602-3880

Abstract

This article studies the implementation of some well-known principles of information work of biological systems in the input unit of the neuroprocessor, including spike coding of information used in models of neural networks of the latest generation.

The development of modern neural network IT gives rise to a number of urgent tasks at the junction of several scientific disciplines. One of them is to create a hardware platform — a neuroprocessor for energy-efficient operation of neural networks. Recently, the development of nanotechnology of the main units of the neuroprocessor relies on combined memristor super-large logical and storage matrices. The matrix topology is built on the principle of maximum integration of programmable links between nodes. This article describes a method for implementing biomorphic neural functionality based on programmable links of a highly integrated 3D logic matrix.

This paper focuses on the problem of achieving energy efficiency of the hardware used to model neural networks. The main part analyzes the known facts of the principles of information transfer and processing in biological systems from the point of view of their implementation in the input unit of the neuroprocessor. The author deals with the scheme of an electronic neuron implemented based on elements of a 3D logical matrix. A pulsed method of encoding input information is presented, which most realistically reflects the principle

Citation: Pisarev A. D. 2019. “Energy efficient biomorphic pulse information coding in electronic neurons for the entrance unit of the neuroprocessor”. Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 5, no 3, pp. 186-212.
DOI: 10.21684/2411-7978-2019-5-3-186-212

of operation of a sensory biological neural system. The model of an electronic neuron for selecting ranges of technological parameters in a real 3D logic matrix scheme is analyzed. The implementation of disjunctively normal forms is shown, using the logic function in the input unit of a neuroprocessor as an example. The results of modeling fragments of electric circuits with memristors of a 3D logical matrix in programming mode are presented.

The author concludes that biomorphic pulse coding of standard digital signals allows achieving a high degree of energy efficiency of the logic elements of the neuroprocessor by reducing the number of valve operations. Energy efficiency makes it possible to overcome the thermal limitation of the scalable technology of three-dimensional layout of elements in memristor crossbars.

Keywords

Nanoelectronics, neural networks, distributed computing, biomorphic neuroprocessor, combined crossbar, memristor.

DOI: 10.21684/2411-7978-2019-5-3-186-212

REFERENCES

1. Barsky A. B. 2004. Neural Networks: Recognition, Management, Decision Making. Moscow: Finance and Statistics. [In Russian]
2. Bloom F., Leiserson A., Hofstadter L. 1988. Brain, Mind, and Behavior. Translated from English. Moscow: Mir. [In Russian]
3. Bobylev A. N., Udovichenko S. Yu., Busygin A. N., Ebrahim A. H. 2019. "Increase of switching range of resistive memristor for realization of a greater number of synaptic states in a neuroprocessor". Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 5, no 2, pp. 124-136. DOI: 10.21684/2411-7978-2019-5-2-124-136 [In Russian]
4. Backus J. 1993. "Is programming free from von Neumann style? Functional style and associated program algebra". In: Lectures by Turing Prize Laureates. Moscow: Mir. [In Russian]
5. Dubrovsky D. I. (ed.). 2013. Global Future 2045. Converged Technologies (NBICS) and Transhumanist Evolution. Moscow: IBA. [In Russian]
6. Gusev A. V. 2017. "Prospects for neural networks and deep machine learning in creating solutions for healthcare". Information Technologies for the Physician, no 3, pp. 92-105. [In Russian]
7. Kalichkin V. K., Pavlova A. I. 2014. "The use of a neural expert system for the classification of erosive lands". Siberian Herald of Agricultural Science, no 6, pp. 5-11. [In Russian]
8. Pisarev A. D., Mayevsky O. V., Busygin A. N., Udovichenko S. Yu. 2018. "A multilayer logic matrix based on a memristor switching cell". Invention patent No 2017122704. Applied 27 June 2017. Issued 10 April 2018. [In Russian]
9. Pisarev A. D., Busygin A. N., Bobylev A. N., Udovichenko S. Yu. 2017. "Combined memristor-diode crossbar as a memory storage base". Tyumen State University Herald.

- Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 3, no 4, pp. 142-149. DOI: 10.21684/2411-7978-2017-3-4-142-149 [In Russian]
10. Pisarev A. D. 2019. "Implementation of discrete cosine transformation in the input block of the memristor neural processor". Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 5, no 1, pp. 147-161. DOI: 10.21684/2411-7978-2019-5-1-147-161 [In Russian]
 11. Pisarev A. D. 2018. "SPICE-Modeling of the Processes of Associative Self Learning and Unconditional Discrimination in the Logic Unit of a Neuroprocessor". Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy, vol. 4, no 3, pp. 132-145. DOI: 10.21684/2411-7978-2018-4-3-132-145 [In Russian]
 12. Trofimova E. A., Mazurov V. D., Gilev D. V. 2017. Neural Networks in Applied Economics. Edited by E. A. Trofimova. Commissioned by the RF Ministry of Education and Science. Yekaterinburg: Ural Federal University. [In Russian]
 13. Udovichenko S. Yu., Pisarev A. D., Busygin A. N., Mayevsky O. V. 2018. "Neuroprocessor based on a combined memristor-diode crossbar". Nanoindustry, no 5 (84), pp. 344-355. DOI: 10.22184/1993-8578.2018.84.5.344.355 [In Russian]
 14. Udovichenko S. Yu., Pisarev A. D., Busygin A. N., Mayevsky O. V. 2017. "3D CMOS-memristor nanotechnology for creating a logical and memory matrix of a neuroprocessor". Nanoindustry, no 5 (76), pp. 26-34. DOI: 10.22184/1993-8578.2017.76.5.26.34 [In Russian]
 15. Biolek D., Di Ventra M., Pershin Y. V. 2013. "Reliable SPICE simulations of memristors, memcapacitors and meminductors". Radioengineering, vol. 22, no 4, pp. 945-968.
 16. Bobylev A. N., Busygin A. N., Pisarev A. D., Udovichenko S. Yu., Filippov V. A. 2017. "Neuromorphic coprocessor prototype based on mixed metal oxide memristors". International Journal of Nanotechnology, vol. 14, no 7/8, pp. 698-704. DOI: 10.1504/IJNT.2017.083444
 17. Bobylev A. N., Udovichenko S. Yu. 2016. "The electrical properties of memristor devices $\text{TiN}/\text{Ti}_x\text{Al}_{1-x}\text{O}_y/\text{TiN}$ produced by magnetron sputtering". Russian Microelectronics, vol. 45, no 6, pp. 396-401. DOI: 10.1134/S1063739716060020
 18. Chen Y.-H., Krishna T., Emer J. S., Sze V. 2017. "Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks". IEEE Journal of Solid-State Circuits, vol. 52, no 1, pp. 127-138. DOI: 10.1109/JSSC.2016.2616357
 19. Filippov V. A., Bobylev A. N., Busygin A. N., Pisarev A. D., Udovichenko S. Yu. 2019. "A biomorphic neuron model and principles of designing a neural network with memristor synapses for a biomorphic neuroprocessor". Neural Computing and Applications. Published online. DOI: 10.1007/s00521-019-04383-7
 20. Gardner B, Grüning A. 2016. "Supervised learning in spiking neural networks for precise temporal encoding". PLoS ONE, vol. 11, no 8, art. e0161335, pp. 1-28. DOI: 10.1371/journal.pone.0161335
 21. Gollisch T., Meister M. 2008. "Rapid neural coding in the retina with relative spike latencies". Science, vol. 319, no 5866, pp. 1108-1111. DOI: 10.1126/science.1149639
 22. Gütig R. 2014. "To spike, or when to spike?". Current Opinion in Neurobiology, vol. 25, pp. 134-139. DOI: 10.1016/j.conb.2014.01.004
 23. Hadiyawardan, Budiman F., Hernowo D. G. O., Pandey R. R., Tanaka H. 2018. "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, vol. 52, no 3S2, art. 03EA06. DOI: 10.7567/JJAP.57.03EA06

24. Hodgkin A. L., Huxley A. F. 1952. "A quantitative description of membrane current and its application to conduction and excitation in nerve". *The Journal of Physiology*, vol. 117, no 4, pp. 500-544. DOI: 10.1113/jphysiol.1952.sp004764
25. Johansson R. S, Birznieks I. 2004. "First spikes in ensembles of human tactile afferents code complex spatial fingertip events". *Nature Neuroscience*, vol. 7, no 2, pp. 170-177. DOI: 10.1038/nn1177
26. Kasinski A., Ponulak F. 2006. "Comparison of supervised learning methods for spike time coding in spiking neural networks". *International Journal of Applied Mathematics and Computer Science*, vol. 16, no 1, pp. 101-113.
27. Larkum M. E., Zhu J. J., Sakmann B. 2001. "Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons". *The Journal of Physiology*, vol. 533, no 2, pp. 447-466. DOI: 10.1111/j.1469-7793.2001.0447a.x
28. Mainen Z. F, Sejnowski T. J. 1995. "Reliability of spike timing in neocortical neurons". *Science*, vol. 268, no 5216, pp. 1503-1506. DOI: 10.1126/science.7770778
29. Mohemmed A., Schliebs S., Matsuda S., Kasabov N. 2012. "SPAN: spike pattern association neuron for learning spatio-temporal spike patterns". *International Journal of Neural Systems*, vol. 22, no 04, art. 1250012. DOI: 10.1142/S0129065712500128
30. Pavlova A. I., Morozov D. A. 2017. "The use of neural networks in forecasting the load of the call-center". *International journal of advanced studies*, vol. 7, no 2-2, pp. 108-112. DOI: 10.12731/2227-930X-2017-2-2-108-112
31. Peng C.-S., Chang W.-Y., Lee Y.-H., Lin M.-H., Chen F., Tsai M.-J. 2012. "Improvement of resistive switching stability of HfO₂ films with Al doping by atomic layer deposition". *Electrochemical and Solid-State Letters*, vol. 15, no 4, pp. H88-H90. DOI: 10.1149/2.011204esl
32. Pisarev A. D., Busygin A. N., Udovichenko S. Yu., Maevsky O. V. 2018. "3D memory matrix based on a composite memristor-diode crossbar for a neuromorphic processor". *Microelectronic Engineering*, vol. 198, pp. 1-7. DOI: 10.1016/j.mee.2018.06.008
33. Reich D. S, Victor J. D, Knight B. W, Ozaki T., Kaplan E. 1997. "Response variability and timing precision of neuronal spike trains in vivo". *Journal of Neurophysiology*, vol. 77, no 5, pp. 2836-2841. DOI: 10.1152/jn.1997.77.5.2836
34. Takeuchi T., Duzkiewicz A. J., Morris R. G. M. 2014. "The synaptic plasticity and memory hypothesis: encoding, storage and persistence". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 369, no 1633. DOI: 10.1098/rstb.2013.0288
35. Uzzell V. J., Chichilnisky E. J. 2004. "Precision of spike trains in primate retinal ganglion cells". *Journal of Neurophysiology*, vol. 92, no 2, pp. 780-789. DOI: 10.1152/jn.01171.2003
36. Van Rullen R., Guyonneau R., Thorpe S. J. 2005. "Spike times make sense". *Trends in Neurosciences*, vol. 28, no 1, pp. 1-4. DOI: 10.1016/j.tins.2004.10.010
37. Yu Q., Tang H., Tan K. C, Li H. 2013. "Precise-spike-driven synaptic plasticity: Learning hetero-association of spatiotemporal spike patterns". *PLoS ONE*, vol. 8, no 11, art. e78318. DOI: 10.1371/journal.pone.0078318