

На правах рукописи

Теплов Георгий Сергеевич

**РАЗРАБОТКА МОДЕЛИ ИСКУССТВЕННОГО НЕЙРОНА С
ДИНАМИЧЕСКОЙ ФУНКЦИЕЙ АКТИВАЦИИ НА БАЗЕ
МЕМРИСТИВНЫХ КОМПОНЕНТОВ**

Специальность 05.27.01 –
твердотельная электроника, радиоэлектронные компоненты,
микро- и наноэлектроника на квантовых эффектах

Автореферат диссертации
на соискание ученой степени
кандидата физико-математических наук

Москва – 2018

Работа выполнена в Акционерном обществе «Научно-исследовательский институт молекулярной электроники» (АО «НИИМЭ») и в федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (государственный университет)» (МФТИ)

Научный руководитель: Горнев Евгений Сергеевич
д-р тех. наук, АО «НИИМЭ», заместитель
руководителя приоритетного технологического
направления по электронным технологиям

Научный консультант: Матюшкин Игорь Валерьевич
к-т физ.-мат. наук, АО «НИИМЭ», старший
научный сотрудник отдела 24

Официальные оппоненты: Яковлев Виктор Борисович,
д-р физ.-мат. наук, профессор, профессор РАН,
институт нанотехнологий микроэлектроники
РАН, начальник отдела

Ревизников Дмитрий Леонидович,
д-р физ.-мат. наук, профессор, Московский
авиационный институт (национальный
исследовательский университет), факультет
«Информационные технологии и прикладная
математика»

Ведущая организация: Федеральное государственное бюджетное
учреждение «Национальный исследовательский
центр «Курчатовский институт»

Защита состоится «___» _____ 2018 года в _____ часов на заседании
диссертационного совета Д 002.081.01 при Федеральном государственном
бюджетном учреждении науки Институте проблем технологии микроэлектроники
и особочистых материалов Российской академии наук по адресу: 142432,
Московская область, г. Черноголовка, ул. Академика Осипьяна, д. 6, ИПТМ РАН.

С диссертацией можно ознакомиться в библиотеке ИПТМ РАН и на сайте
http://www.iptm.ru/tes_council/diss.ru.html

Автореферат разослан «__» _____ г.

Учёный секретарь
диссертационного совета Д 002.081.01,
кандидат физико-математических наук

А.В. Иржак

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Уменьшение минимального топологического размера при производстве микросхем, появление новых электронных компонентов – мемристоров, phase change memory (PCM) – элементов памяти, spin-transfer torque magnetic random access memory (STT-MRAM) и spin orbit torque (SOT) компонентов, развитие методов и средств распараллеливания вычислений, привнесли в аппаратные реализации искусственных нейронных сетей (ИНС) и искусственных нейронов новые факторы, приводящие к необходимости пересмотра существующей структурной схемы имплементации нейронов и нейроморфных систем.

Существующие модельные представления искусственных нейронов, несмотря на подробное изучение различных концепций описания искусственного нейрона, не учитывают возможность динамического изменения функции активации в процессе вычислений или обучения, что позволяет сделать вывод об актуальности и научной новизне исследуемого вопроса.

Исследование мемристивных компонентов в качестве элементов искусственных нейронов требует учета электрофизических особенностей его имплементации в рамках структурно–функциональной схемы одиночного нейрона, в частности, разброса (девиации) выходного сигнала мемристора, что позволит более корректно проектировать последующие схемы обработки сигналов. Существующие представления мемристивных элементов, прежде всего на базе тонких пленок нестехиометрических оксидов Ti, Hf, Si, Ta и других материалов, для компактного моделирования их характеристик средствами САПР не позволяют производить указанный учет. Таким образом, актуальна задача построения компактной модели мемристора, реализованной на языке высокого уровня.

Практическая актуальность темы обусловлена получением математического описания искусственного нейрона, учитывающего вариативность функций агрегации входного сигнала и вариативность функций активации нейрона, в рамках единой структурно-функциональной схемы аппаратной реализации нейрона, а также, возможностью применения разрабатываемых модельных представлений мемристора в качестве библиотечного элемента САПР при проектировании устройств микроэлектроники.

Степень разработанности темы

Параметры аппаратной реализации искусственных нейронов и степень их влияния на вычислительные возможности сети в целом, довольно подробно изучены в мировой и отечественной литературе. Все существующие модели нейронов можно разделить на два обширных класса: модели с учетом

предыдущего состояния нейрона и модели без учета предыдущего состояния. К моделям второго класса можно отнести модели нейрона, предложенные в работах, McCulloch W.S., Pitts W.A. (1943), формальный нейрон Розенблатта (1957); алгебраический $\Sigma\Pi$ -нейрон, подробно рассматриваемый в работах З.М. Шизбухова, Е.В. Бодянского и др (1990/2000); мультипликативный нейрон Koch С. и др. (1992) и многие другие модели. К моделям нейронов, учитывающим свои предыдущие состояния, можно отнести модели integrate-and-fire (1907), спайковые модели нейронов, модель Ходжкина-Хаксли (1952), модель Ижекевича (нач. 80-х гг.) и многие другие. Несмотря на столь широкий спектр типов функций активации и моделей искусственных нейронов, включая гибридные и модели с памятью (синтез с конечными автоматами), модель искусственного нейрона с динамической функцией активации, то есть оперативно изменяющимся в процессе обучения или работы типом функции активации, разработана и рассмотрена не была.

Построение нейроморфных устройств на основе мемристоров, в качестве элемента синапсов, активно изучается в работах коллектива возглавляемого Andrei Dinu, работах И. И. Маранчука, а так же исследованиях коллективов В.А. Демина, Д. Негрова, и многих других. Применение мемристивных элементов для гибридной модели клеточного автомата и нейронной сети (Cellular Neural Network) предложил Chua L.O.(в конце 80-х), а применение резистивной памяти для традиционных вычислительных архитектур – коллектив Strukova D.V. (2008). Вопросы многоуровневой модели синапса рассматриваются в работах А.В. Зенкевича и многих других.

Цели и задачи

Цель работы – разработка функциональной модели искусственного нейрона позволяющего производить выбор типа функции активации в процессе обучения или работы, его структурной схемы для аппаратной реализации с применением мемристивных компонентов.

Для достижения поставленной цели в работе поставлены и решены следующие научные задачи:

- Синтез абстракций формального нейрона и конечного автомата (модель конечного автомата абстрактного нейрона – КААН) и ее спецификация для мультипликативной и аддитивной функции групповой обработки входных сигналов.
- Разработка и описание высокоуровневой функциональной модели искусственного нейрона с динамической функцией активации.

- Разработка модельных представлений мемристора средствами САПР Cadence на языке высокого уровня описания аппаратуры Verilog-A.
- Анализ разработанных модельных представлений мемристора с помощью САПР Cadence.
- Разработка описания нейрона с динамической функцией активации в виде структурной схемы с учетом физических особенностей мемристивных элементов и с возможностью задания множества дискретных состояний.

Объектом исследований является искусственный нейрон, использующий регистры памяти и мемристивные элементы при аппаратной реализации. Предметом исследований является модель нейрона с динамической функцией активации: математическая и формализованная на языке высокого уровня, включающая элементы схемотехнического описания, а так же связанное с ней модельное представление мемристивного элемента средствами САПР Cadence. Для решения поставленных в работе задач использовались методы, основанные на теории множеств, теории автоматов, схемотехнике, теории алгоритмов и прикладного программирования, а так же стандартные методы схемотехнического моделирования САПР Cadence.

Научная новизна

1. Впервые предложена модель нейрона с динамической функцией активации отличающейся тем, что производится выбор функции активации, переключаемой либо в процессе функционирования нейрона, либо в процессе обучения. Показано влияние параметров модели (мощность множества определения функции активации, мощность множества весовых коэффициентов синапсов, мощность алфавита входных и выходных сигналов) на вычислительную мощность искусственного нейрона.

2. Предложено обобщение в рамках предлагаемой математической модели конечного автомата абстрактного нейрона с динамической функцией активации, агрегационных функций математических моделей суммирующего и мультипликативного нейронов. Идея синтеза основана на задании равноможных множеств определения функций активации. Построение эквивалентных моделей реализуется путем подбора элементов множеств значений весовых коэффициентов с последующим (для обеих моделей нейронов) взаимным повторением порядка и следования элементов множества значений при задании функций активации.

3. Предложена математическая модель с неэквидистантным следованием уровней значений весов синапса, позволяющая в предельном случае увеличить

область определения агрегационной функции искусственного нейрона до $I \cdot N^W$, где N — количество синапсов, W — мощность множества значений весовых коэффициентов, I — мощность множества значений входных сигналов.

4. Установлено соответствие между физическими параметрами известных мемристивных компонентов (вольт-амперные характеристики процессов переключения, механизм переключения и другие) и формальными параметрами Verilog-A описания в среде САПР Cadence. Модель мемристора позволяет произвести описание множественности состояний проводимости с учетом девиации следующих параметров: напряжений порогов переключения, высокорезистивного и низкорезистивного состояний, количества циклов переключения. Множественность состояний проводимости, в отличие от существующих моделей, реализуется с учетом отклонений параметров и без необходимости отдельного, ограниченного по количеству и преднамеренно задаваемого параметрами модели описания промежуточных состояний проводимости.

5. Предложена структурная схема реализации искусственного нейрона, включающая два блока LUT (LUT – таблица значений функции) и сдвиговый регистр. Преимущество схемы заключается в учете амплитуд агрегированного сигнала в течение всего времени активации в дополнение к учету обобщенного уровня возбуждения нейрона.

Практическая значимость работы

- Предложенная модель конечного автомата абстрактного нейрона с динамической функцией активации (КААН) позволяет реализовать набор искусственных нейронов с различными функциями активации, что при аппаратной реализации КААН гарантирует относительную универсальность технического решения и широкий набор возможностей при проектировании сети.

- Представленное описание на языке Verilog-A позволяет использовать мемристор в качестве стандартного элемента библиотеки электронных компонентов САПР Cadence, что в свою очередь необходимо для проектирования микросхем с нейроморфной структурой. Предложенное описание может быть модифицировано в зависимости от конкретных параметров создаваемых структур.

- Результаты моделирования демонстрируют предпочтительность снижения разброса параметров высокопроводящего состояния мемристора и параметров разброса пороговых напряжений для изготавливаемых МДМ (металл-диэлектрик-металл) структур. Разброс параметров в высокопроводящем состоянии будет вносить большие искажения в обрабатываемый сигнал в сравнении с влиянием

разброса в низкопроводящем состоянии. Разброс параметров пороговых напряжений при переключении от цикла к циклу напрямую влияет на точность получаемых промежуточных состояний проводимости мемристора.

- Следствием результатов моделирования с учетом разброса параметров мемристора является приоритет задания множественности состояний путем подачи коротких (не более 15 нс) импульсов с малой амплитудой по напряжению (не более 0.15 В выше порога переключения) перед длительными импульсами с малой амплитудой, либо короткими импульсами с большой амплитудой.

Положения, выносимые на защиту

1. Математическая модель искусственного нейрона с динамической функцией активации в виде схемы конечного автомата абстрактного нейрона, реализующая синтез математических моделей искусственных нейронов с агрегирующими функциями сложения и умножения. Динамика функции активации заключается в оперативном изменении функции активации в процессе обучения или работы нейронной сети.

2. Математическая модель синапсов с нелинейной зависимостью между элементами весовых коэффициентов синапсов, позволяющая увеличить максимальную мощность множества определения функции агрегации до $I \cdot N^W$, где N – количество синапсов, W – мощность множества значений весовых коэффициентов, I – мощность множества значений входных сигналов.

3. Модель мемристора, описанная на языке Verilog-A и позволяющая моделировать промежуточные состояния проводимости с учетом девиаций параметров при переключении: разброса напряжений порогов переключения между состояниями проводимости, разбросов низкорезистивного и высокорезистивного состояний, разброса количества циклов переключения.

4. Структурная схема реализации искусственного нейрона, включающая два блока LUT и сдвиговый регистр, позволяющая учитывать не только общий уровень активности агрегированных входных сигналов, но и влияние амплитуды агрегированных входных сигналов на выходной сигнал для каждого момента времени активации.

Достоверность научных положений, результатов и выводов

Достоверность научных результатов обуславливается применением методов теории множеств, при построении модели КААН, а также классических методов комбинаторики при исследовании нейрона с динамической функцией активации. Обоснованность Verilog-A описания следует из высокого уровня результатов, достигнутого в области мемристивных структур.

Достоверность технических результатов обеспечивается использованием современных методов моделирования с применением САПР и программной верификацией работы алгоритма.

Апробация результатов

Результаты работы были представлены на следующих конференциях и семинарах:

- X научно-техническая конференция молодых специалистов «Высокие технологии атомной отрасли. Молодежь в инновационном процессе», Нижний Новгород, 10-12 сентября 2015г.
- «Электроника-2015» Международная научно-техническая конференция, г. Зеленоград, 19-20 ноября 2015г.
- 60-я Научная конференция МФТИ, г. Долгопрудный, 15 октября 2017г.
- Научный семинар «Нейроморфные системы и их реализация» научного совета РАН «Фундаментальные проблемы элементной базы информационно-вычислительных и управляющих систем и материалов для ее создания», г. Зеленоград, 24 сентября 2018.
- 11-ый научно-практический семинар «Математическое моделирование в материаловедении электронных наноструктур», ВЦ РАН имени А.А. Дородницына ФИЦ ИУ РАН, г. Москва, 15 мая.

Исследования были поддержаны грантом РФФИ «Исследование и разработка нейросетевых и клеточно-автоматных технологий в проектировании сверхбольших интегральных схем» № 17-07-00570 А (2017-2018г.).

Публикации

По теме диссертации опубликованы 5 работ в научных журналах и сборниках трудов международных и российских конференций, в том числе 2 работы в рецензируемых журналах, входящих в действующий перечень ВАК.

Личный вклад автора

Все теоретические результаты предлагаемой модели искусственного нейрона получены автором лично, либо в соавторстве при его непосредственном определяющем или весомом участии. Высокоуровневое модельное описание мемристивного элемента на языке Verilog-A и моделирование его работы в среде САПР Cadence, а так же разработка структурной схемы модели КААН на базе мемристивных компонентов выполнена автором лично.

Результаты диссертационной работы, составляющие научную новизну, практическую значимость и выносимые на защиту, получены автором лично под

научным руководством к-та физ.-мат. наук Матюшкина И.В. и д-ра тех. наук Горнева Е.С.

Структура и объём работы

Диссертационная работа состоит из введения, трех глав, заключения, списка терминов и определений, списка сокращений, списка работ, опубликованных по теме диссертации, списка цитируемой литературы из 143 наименований, двух приложений и содержит 156 страниц, в том числе 41 рисунок и 4 таблицы.

СОДЕРЖАНИЕ РАБОТЫ

Введение включает основную сводную информацию по работе. В рамках данной части рукописи формулируются цели и задачи работы, актуальность работы, степень разработанности темы, научная новизна. Раздел так же включает описание практической значимости, перечень положений выносимых на защиту.

Первая глава работы начинается аналитическим обзором математических моделей искусственных нейронов и искусственных нейронных сетей. При рассмотрении моделей искусственных нейронов акцент, смещен к анализу влияния параметров моделей на вычислительную мощность. Базовая концепция представления нейрона в виде блоков учета вклада входного сигнала, блока агрегации сигналов и блока генерации выходного сигнала является классической. Основанные на базовой концепции (рис. 1) модели учитывают:

- Влияние применяемых числовых множеств описания сигналов сети и параметров нейрона на его свойства, рассматриваемое в моделях McCulloch – Pitts, Adaline, нечёткого нейрона и модели комплекснозначного нейрона.
- Подходы к описанию связей нейрона, учитывающие возможность применения весовых коэффициентов на аксонах, исследуемые в работах Grossberg`а;
- Влияние функции агрегации входных сигналов на процесс вычислений, рассматриваемый в моделях суммирующего нейрона, мультипликативного нейрона, агрегирующего нейрона, комбинированного нейрона, адаптивного нейрона и $\Sigma\Pi$ -нейрона.
- Влияние функции активации на процесс вычислений без учета временной динамики процесса активации, учитываемое в моделях WTA-нейронов (winner take all), стохастических нейронов, RBF (radial basis function) нейронов;
- Влияние функции активации на процесс вычислений с учетом временной динамики процесса активации, учитываемое в моделях интегрирующего и интегрирующего с утечками нейронов, а также в модели связывающего нейрона;



Рисунок 1. Базовая концепция искусственного нейрона.

На основе обзора литературы сделан вывод об отсутствии в литературе математического формализма модели искусственного нейрона с динамической функцией активации.

В качестве основных ограничений накладываемых на модель, исходя из условия ее аппаратной реализации, выступают, прежде всего, ограниченность любой вычислительной системы реализуемой аппаратными средствами и дискретность состояний любой реализуемой системы. Дискретность состояний системы напрямую следует из необходимости хранить обрабатываемые данные и соотношения сигнала к шуму. Помимо перечисленных выше ограничений немаловажным с практической точки зрения является требование *универсальности* модели, под которым подразумевается возможность с помощью модели описывать некоторый набор моделей искусственных нейронов.

Исходя из указанных ограничений, автором была предложена математическая модель *конечного автомата абстрактного нейрона* (КААН далее по тексту). Взятие за основу формализма теории автоматов обосновано её широким применением при проектировании микропроцессоров. Изложение модели на языке теории множеств является общепринятым в современной математике.

Модель обобщает два основных типа функций агрегации, применяемых в аппаратных реализациях искусственных нейронов: n -мерные операции сложения [1] и умножения [2].

Как известно, мультипликативные нейроны точнее отражают работу биологических нейронов, чем аддитивные (это относится и к блочным сигма-пи-сигма и пи-сигма-пи нейронам). В алгебре использование двух операций (поле) удобнее, чем использование одной (группа). Альтернативный авторский подход (КААН) не предполагает сложные (составные) модели синапсов для функции агрегации. Возможность применения в рамках авторской модели упрощенных подходов к описанию синапсов напрямую следует из возможности их замены эквивалентной моделью. Необходимое условие замены следующее —

эквивалентная модель должна обладать множеством весовых коэффициентов обеспечивающих при выполнении функции учета весового коэффициента входного сигнала множество неупорядоченных пар равномоощное множеству выходных сигналов синапса исходной модели.

Обобщение мультипликативного и аддитивного типов нейронов в модели абстрактного нейрона возможно при условии равномоощности и упорядоченности множеств определения функции активации искусственного нейрона и при равномоощности области значений данных функций. В рамках работы предлагается определение эквивалентности КААН, базирующееся на определении эквивалентности автоматов (аналогичное определению равенства функций) и включающее в себя следующие условия:

- равенство входных и выходных множеств;
- при подаче на вход любого из эквивалентных искусственных нейронов символа из множества их входных символов выходные символы на выходе нейронов равны.

Приводится доказательство возможности построения эквивалентных моделей искусственных нейронов, параметры которых, все кроме количества синапсов, определены на поле \mathbb{R} . Для моделей, основывающихся на конечных множествах, продемонстрирован метод построения модели аддитивного нейрона, эквивалентной модели мультипликативного нейрона.

При построении модели КААН за основу был взят автомат Мура. По сравнению с автоматом Мили, простота записи позволяет разграничить функцию агрегации и функцию активации нейрона.

КААН описывается набором конечных множеств $(N, E, W, Q, \Delta, \Lambda, T)$, где:

N — множество индексов входов, включая управляющие входы $N \in N$

E — алфавит, включающий алфавит всех возможных входных символов E_{input} и алфавит всех возможных выходных значений E_{output} .

W — линейно упорядоченное конечное множество всех возможных значений весовых коэффициентов информационных входов.

Q — область определения функций активации.

Δ — функция агрегации, зависящая от операции учета весового коэффициента синапса.

Λ — множество функций активации, реализуемых на данном нейроне.

T — множество команд.

На основе E_{input} формируются подмножества входных информационных сигналов E_{in} и множества управляющих сигналов E_r , причём $E_{input} = E_{in} \cup E_r$. Множества символов на различных входах образуют алфавит входных векторов нейрона I_{in} . Алфавит выходных символов имеет вид

$O = \{o_i | o_i = (e_{out}, e_r), e_{out} \in E_{out}, e_r \in E_r, E_{out} \cup E_r = E_{output}\}$, где E_{out} — множество символов информационных выходных сигналов.

Одна из множества функция активации КААН определяется: $\lambda_i \in \Lambda, i \in N$ и имеет следующее описание $\lambda_i: Q \rightarrow O$ и $o^t = \lambda_i(q^t)^t$ для нейрона в синхронной сети. Нейрон асинхронной сети имеет описание $\lambda_i: Q \rightarrow O$ и $o^{t_A} = \lambda_i(q^t)^{t_A}$, где t_A — время процесса активации нейрона, t — время.

Всевозможные комбинации весовых коэффициентов на всех входах КААН представлены в виде множества векторов V , где v_i — элемент этого множества (вектор, одна из возможных комбинаций). Функция агрегации Δ является заданной на множестве декартового произведения I_{in} и V и имеет вид: $\Delta: I_{in} \times V \rightarrow Q$. Динамика во времени описывается: $q^{t+1} = \Delta_{i=1}^n (\delta(v_i^t, i_i^t))^t$ — для синхронных ИНС и $q^{t+1} = \Delta_{i=1}^n (\delta(v_i^t, i_i^t), q^t)^t$ — для асинхронных ИНС. Функция изменения параметров КААН: $\tau_i \in T, i \in N$ в наиболее общем виде описывается как: $\tau_i: E_r \rightarrow V \times \Delta \times \Lambda \times T$, соответственно динамика изменения параметров имеет вид: $(v_i^{t+1}, \Delta^{t+1}, \lambda_i^{t+1}, \tau_i^{t+1}) = \tau_j(i_i)^t, \tau_i^{t+1} \in T, \tau_j \in T$.

Из описания модели следует, что мощность множества определения функции агрегации нейрона напрямую зависит от мощности множества входных сигналов и мощности множества весовых коэффициентов. Описание моделей в основном производится путем использования в качестве операции учета весового коэффициента операции умножения, а в качестве функции агрегации применяются операции умножения или сложения. Из коммутативности обеих операций следует обратная зависимость множества определения функции агрегации от пересечения множества входных сигналов и множества весовых коэффициентов. Другим фактором, приводящим к сужению области определения функции агрегации искусственного нейрона, является пересечение между множествами значений весовых коэффициентов, принимаемых синапсами. Третьим фактором является пересечение множеств входных значений на синапсах, так как множества сигналов на разных входах могут иметь не пустое множество пересечений.

При задании нелинейной зависимости между элементами значений весовых коэффициентов, мощность области определения функции агрегации нейрона растет до $I * N^W$, где N — количество синапсов, W — мощность множества значений весовых коэффициентов, I — мощность множества значений входных сигналов.

Множество Q линейно упорядочено и является областью определений для функций из Λ , и одновременно областью значений для функций Δ .

В целях оптимизации, операция Δ может быть заменена на функцию, состоящую из функции агрегации (сложение или умножение) и взятие модуля от

результата данной операции. В случае задания почти периодической функции мощность множества Q должна соответствовать мощности множества результата функции агрегации.

Соответствие между КААН и Автоматом Мура показано в Таблице №1

Таблица 1. Соответствие между автоматами Мура и КААН

Объект соответствия	Автомат Мура $A =$ $(S, S_0, X, Y, \delta, \lambda)$	КААН $FAAN = (N, E, W, Q, \Delta, \Lambda, T)$
Множество внутренних состояний	S	$V \times \Delta \times \Lambda \times T$
Алфавит входных символов	X	I_{in}
Алфавит выходных символов	Y	O
Начальное состояние	S_0	$\Delta_{i=1}^n \left(\delta(v_i^{t=0}, \emptyset)^{t=0} \right)^{t=0}$ $\tau_j(\emptyset)^{t=0}$
Функция переходов в новое состояние	$\Delta : S \times X \rightarrow S$	$\Delta : I_{in} \times V \rightarrow Q$ $\tau_i : E_r \rightarrow V \times \Delta \times \Lambda \times T$
Функция вывода символов	$\lambda : S \rightarrow Y$	$\lambda_i : Q \rightarrow O$
Динамика переходов	$s^{t+1} = \Delta(s^t, x^t)$	$q^{t+1} = \Delta_{i=1}^n (\delta(v_i^t, i_i^t)^t)^t$ $(v_i^{t+1}, \Delta^{t+1}, \lambda_i^{t+1}, \tau_i^{t+1}) = \tau_j(i_i)^t$
Динамика вывода символов	$y^t = \lambda(s^t)$	$o^t = \lambda_i(q^t)^{tA}$

Предлагаемая модель КААН требует модифицировать алгоритмы обучения для изменения активационной функции КААН и построить алгоритм, автоматически генерирующий сеть с оптимальными активационными функциями нейронов. Появляется также возможность задания алгоритма работы сети с динамически изменяющимися функциями активации каждого нейрона сети. Введённое множество T удобно для формализации внешних инструкций по такому изменению, что открывает широкие возможности для настройки работы сети.

Во второй главе работы производится анализ современных подходов к построению искусственных нейронных сетей на аппаратной основе. В рамках литературного обзора рассматриваются существующие концепции и методы

построения архитектур. Особое внимание уделяется применению современной компонентной базы в качестве элементов искусственного нейрона.

В качестве технических решений при создании нейрочипов, нейроблоков и нейросопроцессоров архитектурно воплощающих ИНС можно выделить следующие три основных подхода:

1. Применение различных видов резистивной памяти в качестве синапсов. Характерными примерами могут выступать SOT-RAM, MRAM, ReRAM (Resistive random access memory), PCM.

2. Разработка специализированного вычислителя на базе ПЛИС.

3. Применение при проектировании чипов цифровой электронной компонентной базы, с архитектурой MIMD или SIMD для ускорения вычислений.

По результатам аналитического обзора сделаны выводы об эффективности реализации ИНС аппаратными средствами в виде набора нейросинаптических ядер на кристалле, либо блоком в виде одного или нескольких нейросинаптических ядер в качестве сопроцессора или специализированного вычислителя. Из всего набора мемристивных элементов наиболее подходящим для применения в искусственных нейронах выступает биполярный мемристор [3]. Критериями эффективности применения мемристоров служат площадь, используемая на кристалле для его реализации и количество получаемых на одном элементе дискретных состояний. В сравнении с другими хорошо масштабируемыми компонентами наноэлектроники ключевым преимуществом мемристоров выступает множественность состояний проводимости, что при задании весового коэффициента позволит сэкономить площадь на кристалле, в сравнении с тем же количеством состояний для PCM, FeRAM (Ferroelectric random access memory), или SOT структур.

Применение униполярных мемристоров (рис. 2а) в качестве элемента кроссбара существенно усложнит процесс переключения и настройки связи искусственного нейрона. Проблемы возникают в связи с большим разбросом параметров на элементах кроссбара, результатом которого могут являться: неполное (не до требуемого уровня проводимости) переключение элементов, частичное увеличение или уменьшение электропроводимости соседних элементов на кроссбаре. Применение в качестве элемента синапса на кроссбаре биполярного мемристора (рис. 2б) позволит избежать таких негативных эффектов, что достигается за счет разной полярности операций переключения между состояниями проводимости.

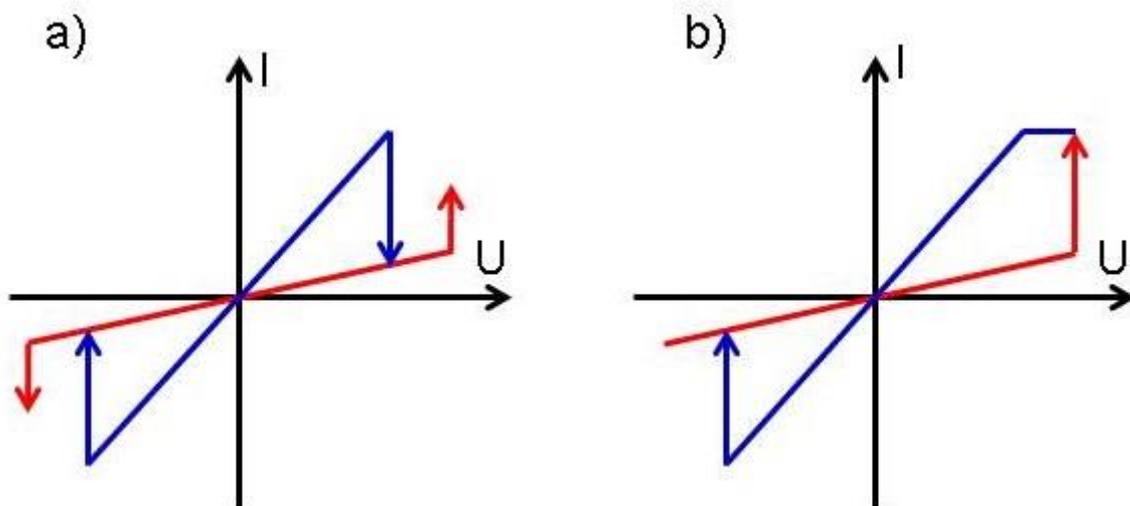


Рисунок 2. а) ВАХ униполярного мемристора б) ВАХ биполярного мемристора.

Предлагаемая автором модель описания мемристора средствами САПР Cadence на языке высокого уровня Verilog-A основывается на модели линейного дрейфа кислородных вакансий с механизмом переключения по напряжению [4]. Отличие заключается в том, что изменение состояния проводимости мемристора зависит от приложенного напряжения, а не протекающего через структуру заряда, и учитывает девиацию параметров мемристора при переключении между состояниями проводимости. В общем виде зависимости имеют вид:

$$I(t) = V(t)/R(V_c, x, t)$$

$$R(V_c, x, t) = R_s - x(V_c, t)$$

$$\frac{dx}{dt} = a * f_w(x) * \begin{cases} \int_0^{t_s} (V_c - V_{thS}) dt, V_c > V_{thS} \\ \int_0^{t_{RS}} (V_c - V_{thRS}) dt, V_c < V_{thRS} \\ 0, V_{thRS} < V_c < V_{thS} \end{cases}$$

$$f_w(x) = \begin{cases} 1, R_{on} + \Delta R_{on} < R(x, t) < R_{off} + \Delta R_{off} \\ 0, else \end{cases}$$

где R_s – начальное состояние проводимости мемристора после электроформовки; a – подгоночный коэффициент; $V(t)$ – приложенное между контактными площадками напряжение, x – изменение сопротивления мемристора; $R(x, t)$ – текущее состояние сопротивления мемристора; $I(t)$ – выходной ток мемристора; V_c – текущее напряжение на контактных площадках мемристора; t_s – время превышения импульсом порога записи; t_{RS} – время превышения импульсом порога стирания; V_{thS} – текущее значение порога переключения в высокопроводящее состояние, задаваемое случайно по усеченному симметричному распределению Гаусса в границах $\pm\Delta V_{thS}$; V_{thRS} – текущее значение порога переключения в

низкопроводящее состояние, задаваемое случайно по усеченному симметричному распределению Гаусса в границах $\pm\Delta V_{thRS}$; R_{on} – текущее значение высокопроводящего состояния, задаваемое случайно по усеченному симметричному распределению Гаусса в границах $\pm\Delta R_{on}$; R_{off} – текущее значение низкопроводящего состояния, задаваемое по усеченному симметричному распределению Гаусса в границах $\pm\Delta R_{off}$.

При необходимости все параметры модели могут быть перенастроены с учетом конкретных физических реализаций. Например, могут быть заданы различные значения порогов переключения в высокорезистивное и низкорезистивное состояние, также изменению могут быть подвержены значения параметров высокорезистивного и низкорезистивного состояний.

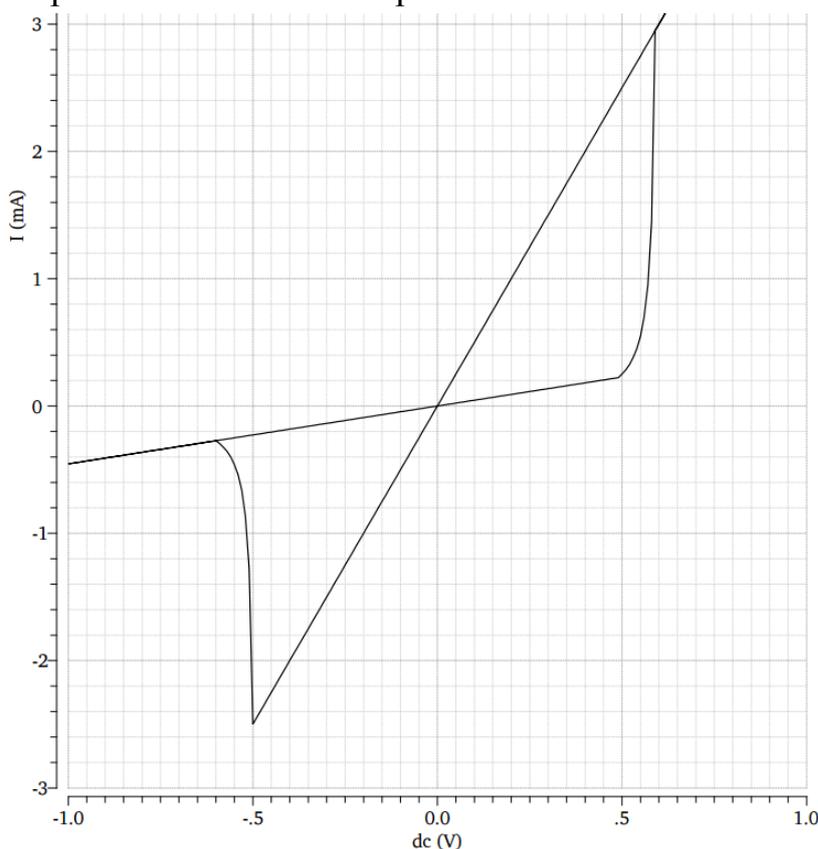


Рисунок 3. Результаты моделирования идеального случая без учета девиаций параметров. $V_{set}=0.5V$, $V_{reset}=-0.5V$. Пример ВАХ гистерезиса модели.

Модель позволяет учитывать количество циклов переключения мемристора, задаваемых напрямую параметром модели, аналогично подходу, применяемому в работе [5]. Отличие от модели, применяемой в указанной работе, заключается в использовании не количества циклов напрямую, а применении переменного значения ресурса переключения мемристора определяемого по формуле:

$$Cyc = NumCyc * (R_{off} + \Delta R_{off} - R_{on} - \Delta R_{on}) * 2$$

где, C_{yc} – значение ресурса на текущей итерации; $NumC_{yc}$ – количество циклов переключения. Изменение значений ресурса переключений описывается выражением:

$$C_{yc}(t + 1) = C_{yc}(t) - x(V_C, t)$$

Указанные особенности описания мемристора позволяют добиться дисперсии выходного сигнала при изменении состояния проводимости мемристора в зависимости от порогов переключений в высокорезистивное и низкорезистивное состояния, а так же учета разброса параметров низкорезистивного (рис. 4) и высокорезистивного (рис. 5) состояний при переключении.

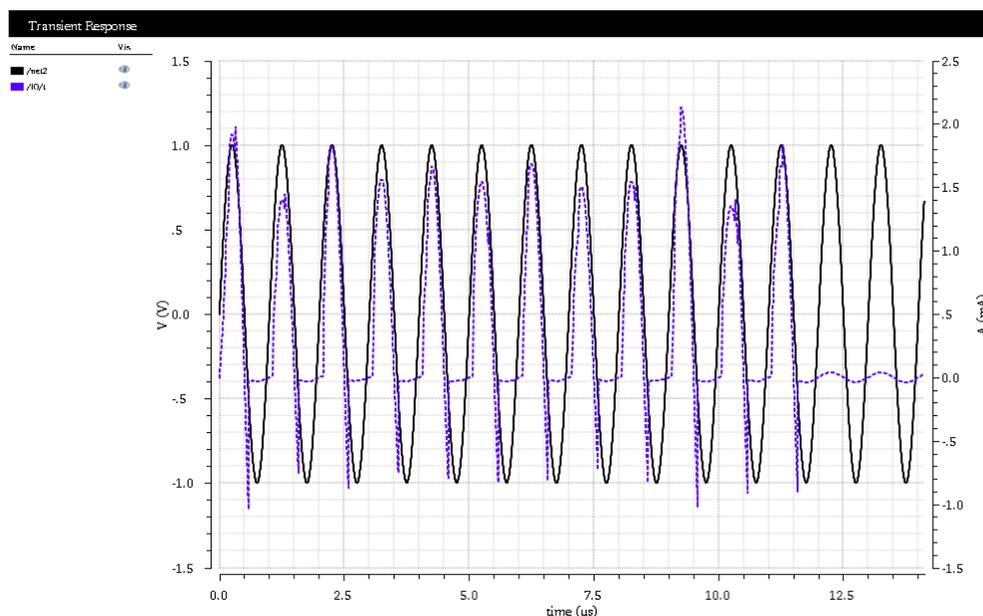


Рисунок 4. Девиация параметров низкорезистивного состояния при циклировании. Черная линия напряжение импульса, синяя линия ток на выходе. $V_{set}=0.5B\pm 0.1$, $V_{reset}=-0.5B\pm 0.1$, $R_{on}=1k\Omega\pm 100\Omega$, $R_{off}=22k\Omega\pm 1k\Omega$

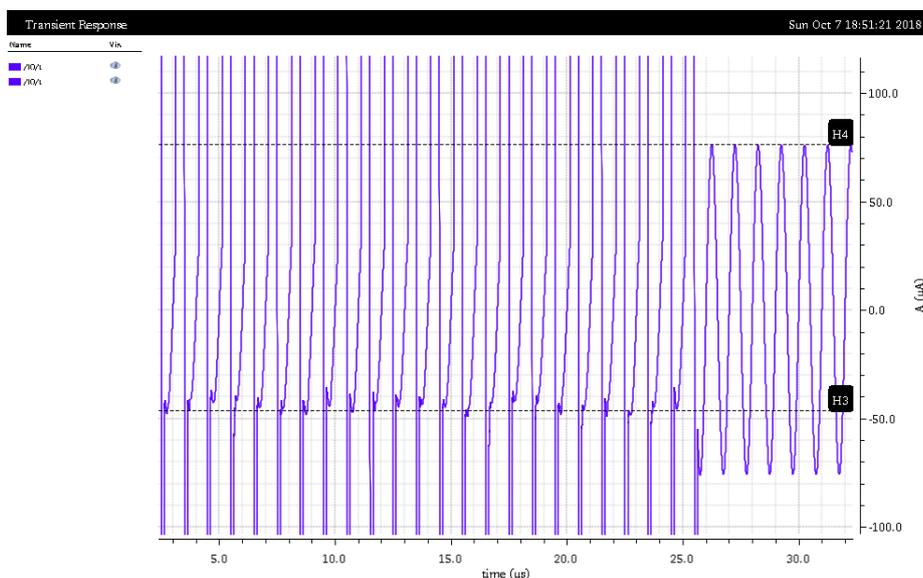


Рисунок 5. Девиация параметров высокорезистивного состояния. Синяя линия ток на выходе. $V_{set}=0.5V\pm 0.1$, $V_{reset}=-0.5V\pm 0.1$, $R_{on}=1k\Omega\pm 100\Omega$, $R_{off}=22k\Omega\pm 1k\Omega$

Введение переменной ресурса переключения позволяет моделировать случаи отказа мемристоров в высокопроводящем (рис. 6а), низкопроводящем (рис. 6б) и промежуточном состоянии проводимости, а также учитывать различное время отказа компонентов в зависимости от разброса параметров.

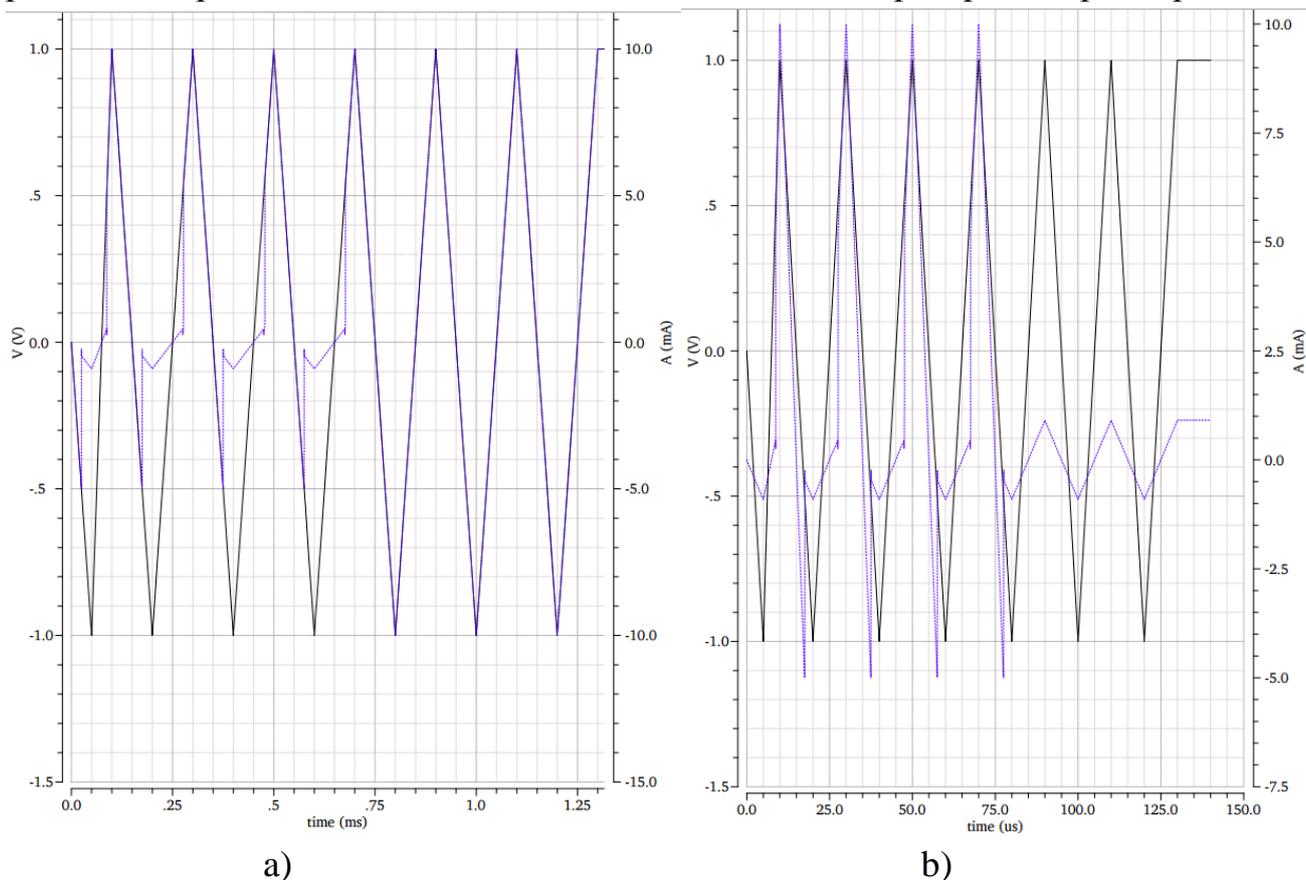


Рисунок 6. а) отказ в высокопроводящем состоянии; б) отказ в низкопроводящем состоянии. $V_{set}=0.5V\pm 0.1$, $V_{reset}=-0.5V\pm 0.1$

Прямым следствием из результатов моделирования является предпочтительность снижения разброса параметров высокопроводящего состояния и параметров разброса пороговых напряжений для изготавливаемых МДМ структур, так как данные разбросы будут напрямую влиять на корректную обработку информационных сигналов и проведение операций переключения. Немаловажным следствием из результатов моделирования является приоритет задания множественности состояний путем подачи коротких (не более 15 нс) импульсов с малой амплитудой по напряжению (не более 0.15В выше порога) перед длительными уровнями с малой амплитудой, либо короткими уровнями с большой амплитудой.

Третья глава посвящена спецификации модели КААН до структурной схемы с учетом применения мемристивных компонентов. Существующие технические решения имплементации функции активации аппаратными средствами искусственного нейрона, при необходимости задания функции активации произвольного типа и ускорения процесса вычислений используют LUT [6]. Указанный метод применяется не только для срабатывающих по совпадению нейронов, но и для интегрирующих, интегрирующих с утечками нейронов и связывающих нейронов [7].

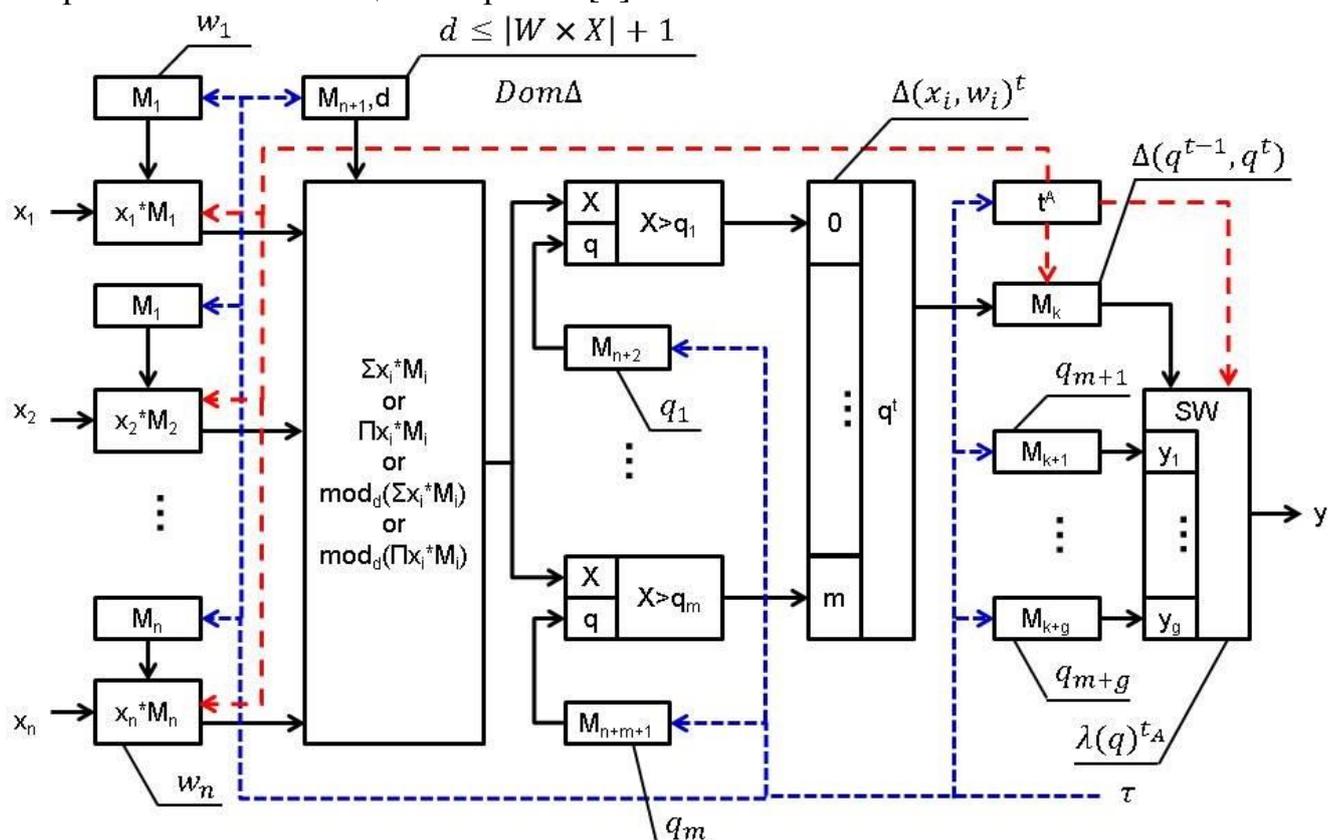


Рисунок 7. Уточненная структурная схема КААН с динамической функцией активации. M – регистры хранения данных, $x * M$ – блок умножения на весовой коэффициент, Σ or Π or $\text{mod}_d(\Sigma$ or $\Pi)$ – блок агрегации входных сигналов, $X > q$ – блоки сравнения с диапазоном значений LUT1, $\Delta(x, w)$ – блок генерации выходного сигнала LUT1, $\Delta(q^{t-1}, q^t)$ – сдвиговый регистр, SW – блок генерации выходного сигнала LUT2, t^A – блок таймера.

Использование, в целях учета текущего состояния активации, в интегрирующих и связывающем нейронах блока суммирования с накоплением позволяет учитывать только обобщенный уровень активности. Последовательность агрегированных сигналов обобщается до уровня общей активности за период активации, без учета последовательности и амплитуды агрегированных в момент времени входных сигналов. Уточненная структурная

схема аппаратной реализации представлена на рис. 7. Предложенная модель КААН с динамической функцией активации не имеет указанного недостатка за счет применения двух блоков LUT и регистра M_k . Указанная опция реализуется за счет применения сдвигового регистра. Выходной сигнал LUT1, фактически, является частью адреса, подаваемого в сдвиговый регистр по сигналу таймера. Работа нейрона в режиме активации осуществляет постепенное формирование адреса ячейки памяти в LUT2, что позволяет учесть не только общий уровень активности на входах нейрона, но и амплитуду уровней агрегированных сигналов в каждый момент времени.

Стоит отметить, что применение мемристоров в обработке сигналов КААН, сопровождается переходом к аналоговой схемотехнике при вычислении выходных сигналов от мемристоров. Активационная функция искусственного нейрона представляется к реализации средствами стандартной КМОП технологии. Уточненная структурная схема КААН с динамической функцией активации на базе мемристивных компонентов представлена на рис. 8.

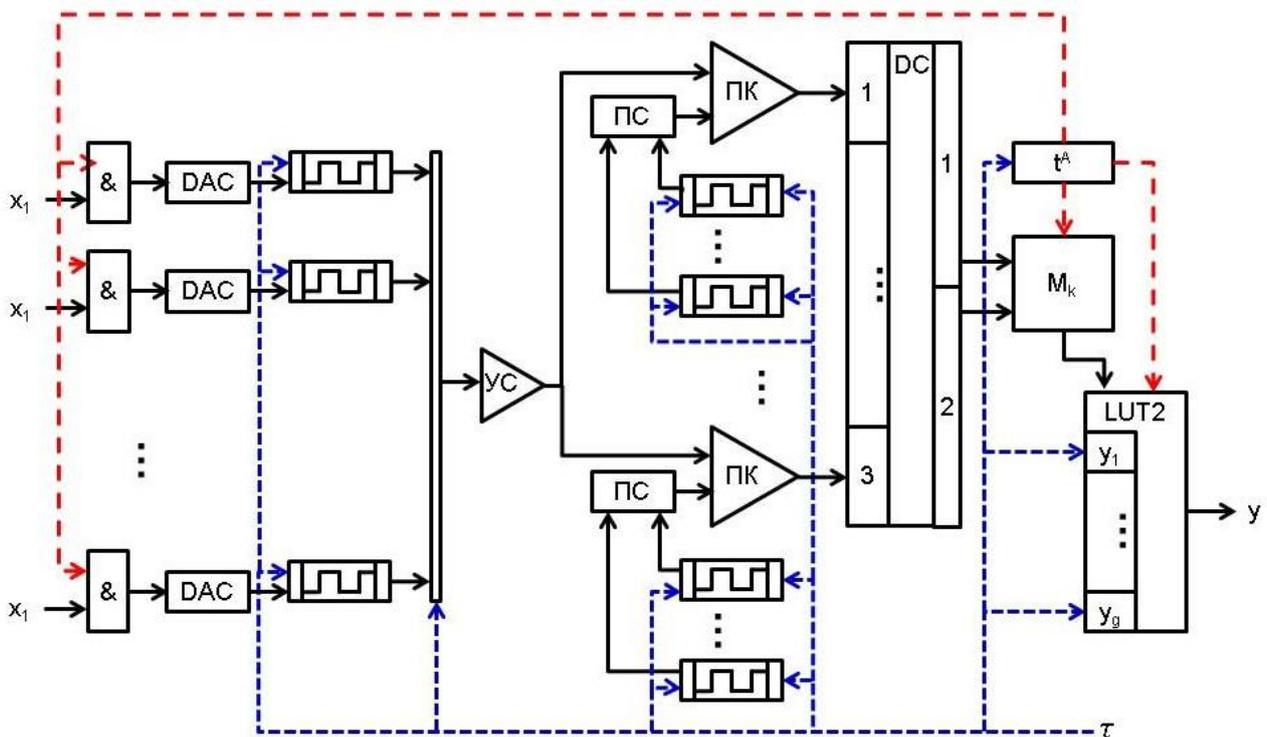


Рисунок 8. Уточненная структурная схема КААН с динамической функцией активации на базе мемристивных компонентов. DAC – цифроаналоговый преобразователь сигнала, УС – блок усиления сигнала, ПК – программируемый компаратор, DC – дешифратор; M_k – сдвиговый регистр, LUT2 – блок генерации выходного сигнала, t^A – блок таймера. Синяя пунктирная линия – линии команд. Красная пунктирная линия – линии сигналов таймера. Черная сплошная линия – линия передачи информационного сигнала.

С учетом применения мемристора с биполярным механизмом переключения по превышению порога напряжения блок учета вклада входного сигнала (рис.8) конкретизируется до следующих блоков: блок контроля учета сигнала по команде таймера, представляется элементом И; блок ЦАП входного сигнала; мемристор. Общее предлагаемое количество синапсов искусственного нейрона 64. Предполагаемое количество дискретных состояний реализуемых на мемристоре 8. Входной сигнал представляется однобитным. В качестве блока агрегации применяется нижняя шина кроссбара с присоединенными к ней синапсами искусственного нейрона, осуществляющая сложение токов от мемристоров. Сигнал от блока агрегации поступает на усилитель и далее на три программируемых компаратора с разрядностью кода в 10 бит. Программирование компараторов предполагает использование блока мемристоров для задания требуемого сигнала сравнения. Сигнал от блока мемристоров подается на преобразователь сигналов и далее, после преобразования, передается на инвертирующий вход. Программируемые компараторы генерируют двоичный позиционный код частичного адреса ячейки LUT2 в конце каждого цикла учета текущей активности на входах искусственного нейрона. Двоичный позиционный код поступает на шифратор, где преобразуется в выходной двухбитный сигнал. Шифратор выступает в качестве ответной части блока LUT1, представляющей агрегированный уровень текущей активности на входах нейрона в виде части двоичного адреса ячейки памяти LUT2.

Данная схема позволяет реализовывать на одной аппаратной платформе, без потери вычислительной мощности, программные решения основных арифметических и логических операций для различных форматов данных, что в свою очередь позволяет производить перепрограммирование ИНС в зависимости от текущих актуальных задач и существенно повышает универсальность технического решения.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

1. Произведено математическое описание искусственного нейрона с динамической функцией активации, переключаемой либо в процессе функционирования нейрона, либо в процессе обучения. В рамках модели осуществлено обобщение агрегационных функций, реализуемых операциями сложения и умножения. Модель позволяет применять метод неэквидистантного задания значений весовых коэффициентов синапсов, что в предельном случае увеличить область определения агрегационной функции искусственного нейрона до $I * N^W$, где N — количество синапсов, W — мощность множества значений весовых коэффициентов, I — мощность множества значений входных сигналов.

2. Установлено соответствие между физическими параметрами известных мемристивных компонентов (вольт-амперные характеристики процессов переключения, механизм переключения, и другие) и формальными параметрами описания на языке Verilog-A в среде САПР Cadence, позволяющее использовать мемристор как библиотечный элемент САПР. Предлагаемое описание позволяет произвести описание множественности состояний проводимости с учетом девиации напряжений порога переключения между циклами переключения, произвести учет параметров разброса высокорезистивного и низкорезистивного состояний между циклами переключения, а также учет параметров разброса количества циклов переключения.

3. Продемонстрировано преимущество реализации модели КААН с динамической функцией активации, включающей два блока LUT и сдвиговый регистр, заключающееся в учете амплитуд агрегированного сигнала в каждый момент времени активации в дополнении к учету обобщенного уровня возбуждения нейрона.

4. Из результатов моделирования следует предпочтительность снижения разброса параметров высокопроводящего состояния и параметров разброса пороговых напряжений для изготавливаемых МДМ структур. Разброс параметров в высокопроводящем состоянии будет вносить большие искажения в обрабатываемый сигнал в сравнении с влиянием разброса в низкопроводящем состоянии. Разброс параметров пороговых напряжений при переключении от цикла к циклу напрямую влияет на точность получаемых промежуточных состояний проводимости мемристора.

5. Следствием из результатов моделирования является приоритет задания множественности состояний путем подачи коротких (не более 15 нс) импульсов с малой амплитудой по напряжению (не более 0.15 В выше порога переключения) перед длительными уровнями с малой амплитудой, либо короткими уровнями с большой амплитудой.

Результаты диссертационной работы могут быть положены в основу дальнейших экспериментальных и теоретических исследований мемристивных элементов и нейроморфных архитектур на их основе.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Теплов Г.С., Матюшкин И.В., Горнев Е.С.** Принципы повышения наработки до отказа схем управления на основе клеточно-автоматных и нейроподобных структур //Атомный проект, Информационно-аналитический журнал для специалистов в области атомного машиностроения –2015 – №. 22. – С. 48-49
2. **Матюшкин И.В., Г.С. Теплов, Е.С. Горнев** Особенности реализации

микросхем с клеточно-автоматной архитектурой на основе эффекта резистивного переключения //«Электроника-2015» Тезисы. Международной научно-технической конференции, МИЭТ, г. Зеленоград, 19-20 ноября 2015, стр. 51-52.

3. **Горнев Е.С., Матюшкин И.В., Теплов Г.С.** Анализ концепций неклассического компьютеринга и парадигмы коннекционизма //Электронная техника. Серия 3: Микроэлектроника.– 2015. – №: 2 (158) – С. 45-66.
4. **Stempkovsky A.L., Gavrilov S.V., Matyushkin I.V., Teplov G.S.** On the issue of application of cellular automata and neural networks methods in VLSI design //Optical memory & Neural Networks. – 2016. – Т. 2. – №. 2. – С. 72-78
5. **Горнев Е.С., Теплов Г.С.** Математическая модель конечного автомата абстрактного нейрона и сетей на его основе //Нано- и микросистемная техника. – 2018. – Т. 20. – №. 7. – С. 434-442.

СПИСОК ЦИТИРУЕМОЙ ЛИТЕРАТУРЫ

1. **Rosenblatt F.** The Perceptron: a probabilistic model for information storage and organisation in the brain //Psychological review. – 1958. – Т. 65. – №6. – С. 386.
2. **Koch C., Poggio T.** Multiplying with synapses and neurons //Single Neuron Computation. – 1992. – С. 315-345.
3. **Strukov D.B. et al.** The missing memristor found //nature. – 2008. – Т.453. – №. 7191. – С. 80.
4. **Zeng G. et al.** Polynomial Metamodel integrated Verilog-AMS for memristor-based mixed-signal system design //Circuits and Systems (MWSCAS), 2013 IEEE 56th International Midwest Symposium on. – IEEE, 2013. – С. 916-9194.
5. **Corinto F. Ascoli A.** A boundary condition-based approach to the modeling of memristor nanostructures //IEEE Transaction on Circuits and Systems I: Regular Papers. – 2012. – Т. 59. – №. 11. – С. 2713-2726.
6. **Piazza F., Uncini A., Zenobi M.** Neural networks with digital LUT activation functions //Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of International Joint Conference on. – IEEE, 1993. – Т. 2. – С. 1401-1404.
7. **Merolla P. A. et al.** A million spiking-neuron integrated circuit with a scalable communication network and interface //Science. – 2014. – Т. 345. – №. 6197. – С. 668-673.