

**К.К.АБГАРЯН
Д.Л. РЕВИЗНИКОВ
А.А.ЖУРАВЛЕВ
А.Ю.МОРОЗОВ
Е.С.ГАВРИЛОВ**

**МНОГОМАСШТАБНОЕ
МОДЕЛИРОВАНИЕ
НЕЙРОМОРФНЫХ
СИСТЕМ**

ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ»
РОССИЙСКОЙ АКАДЕМИИ НАУК

**К.К. Абгарян, Д.Л. Ревизников, А.А. Журавлев,
А.Ю. Морозов, Е.С. Гаврилов**

**МНОГОМАСШТАБНОЕ МОДЕЛИРОВАНИЕ
НЕЙРОМОРФНЫХ СИСТЕМ**

Монография

Москва
Издательство МАКС Пресс
2022



*Работа выполнена при финансовой поддержке РФФИ
по гранту №19-29-03051 мк*

Рецензент:

*Харченко В.А. – д-р техн.наук, вед. науч. сотр. отдела
математического моделирования гетерогенных систем
ФГУ «ФИЦ «Информатика и управление» РАН»*

Абгарян К.К., Ревизников Д.Л., Журавлев А.А., Морозов А.Ю., Гаврилов Е.С.

М54 Многомасштабное моделирование нейроморфных систем : Монография / К.К. Абгарян, Д.Л. Ревизников, А.А. Журавлев, А.Ю. Морозов, Е.С. Гаврилов. – Москва : МАКС Пресс, 2022. – 128 с.

ISBN 978-5-317-06886-8

<https://doi.org/10.29003/m3139.978-5-317-06886-8>

Представлены многомасштабные подходы, модели и алгоритмы для проектирования нейроморфных устройств памяти компьютеров нового поколения. Разработанные подходы дают возможность решать задачи, связанные с имитационным моделированием работы нейроморфных сетей в режимах интеллектуального анализа данных и машинного обучения.

Рассмотрены вопросы, связанные с построением вычислительной модели образования/разрушения проводящих каналов в мемристорных элементах, лежащих в основе нейроморфных сетей. Созданы новые алгоритмы моделирования работы нейроморфной сети с учетом стохастических эффектов, а также построены оригинальные методы и средства имитационного моделирования обучения нейроморфной сети.

Представлены основные подходы, примененные при создании программного обеспечения для имитационного моделирования работы нейроморфных сетей на базе интеграционной платформы многомасштабного моделирования, объединяющей информационные потоки на различных масштабных уровнях, включая уровень мемристора – нового элемента резистивной памяти, уровень нейроморфной сети и уровень имитации обучения нейроморфной сети по прецедентам.

Книга предназначена для научных работников, специалистов в области вычислительной электроники, студентов старших курсов и аспирантов технических университетов.

Ключевые слова: математическое моделирование, нейроморфные системы, аналоговые нейронные сети, мемристивные элементы, стохастическая динамика, филамент, многомасштабное моделирование, гибридные программные системы, высокопроизводительные системы.

УДК 519.6.517.958.004.8
ББК 22.2.2218+32.813.5

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
Глава 1. Моделирование образования/разрушения проводящих филаментов в мемристивных элементах.....	8
1.1. Многомасштабная вычислительная модель	8
1.2. Квантово-механическое и молекулярно-динамическое моделирование.....	9
1.3. Монте-Карло моделирование ионной динамики в мемристивных элементах	19
1.4. Вычислительные эксперименты	27
1.5 Выводы	37
Глава 2. Динамические модели мемристора	42
2.1. Представление мемристора динамической системой. Детерминированные модели	43
2.2. Стохастические модели мемристоров	48
2.3. Интервальные модели мемристоров.....	54
Глава 3. Имитационное моделирование аналоговой самообучающейся импульсной нейронной сети	64
3.1. Детерминированная модель нейронной сети.....	65
3.2. Стохастическая модель нейронной сети	82
3.3. Интервальная модель нейронной сети	91
3.4. Параллельная реализация модели нейронной сети.....	93
Глава 4. Аналоговая реализация классических нейронных сетей	97
Глава 5. Применение интеграционной платформы для многомасштабного моделирования нейроморфных систем.....	103

5.1	Принципы формирования информационной поддержки многомасштабного моделирования физических явлений и процессов.....	103
5.2	Интеграционная платформа для моделирования работы нейроморфной сети.....	104
5.3	Выводы	114
	Список источников.....	116

ВВЕДЕНИЕ

В последнее время в области микроэлектроники наблюдается повышенный интерес к элементам энергонезависимой резистивной памяти, выступающих в качестве перспективной элементной базы для создания аналоговых нейронных сетей, имитирующих функции мозга [1–4]. Элементы резистивной памяти и их матрицы могут работать как искусственные синапсы, тогда как логические схемы комплементарных структур металл-оксид-полупроводник (КМОП) могут функционировать как нейроны. Многоуровневые элементы резистивной памяти позволяют увеличить плотность интеграции энергонезависимой памяти, а также проектировать системы с механизмом параллельных вычислений, необходимые для создания основы адаптивных нейроморфных сетей. Основным элементом резистивной памяти (мемристора) является структура металл-диэлектрик-металл (МДМ). Активной средой, как правило, являются оксиды металлов, такие как HfO_x , ZrO_x , TiO_x , TaO_x , AlO_x , а также оксиды кремния SiO_x и нитриды кремния SiN_x . Переключение между различными состояниями мемристивного элемента на основе тонкого оксидного слоя связано с образованием/разрушением филаментных структур, которые обычно представляют собой кластеры кислородных вакансий [5–7]. Кислородные вакансии действуют как ловушки для электронов и создают условия для туннелирования [8]. Широкий диапазон временных масштабов, присущий рассматриваемому классу явлений, определяет актуальность многомасштабного подхода к моделированию [9–13]. Существует два основных класса математических моделей, описывающих динамику переключения резистивности оксидного слоя. Ко первому классу относятся модели, основанные на физических принципах. Модели различного уровня сложности представлены в [14–20]. Наиболее комплексный подход предполагает прямое моделирование процессов генерации/рекомбинации ионов кислорода, транспорта ионов и электронов в оксидном слое. Как правило, для описания динамики ионов в системе используются Монте-

Карло симуляторы [18–20]. Такой подход связан с большими вычислительными затратами, однако позволяет воспроизвести детальную картину эволюции филаментных структур. В настоящей работе реализуется многомасштабная вычислительная схема, позволяющая проводить моделирование на основе первичной информации – данных о химическом составе материала и его кристаллической структуре. Еще одной отличительной чертой работы является организация моделирования динамики ионов методом Монте-Карло на реальной кристаллической решетке. В главе 1 дается описание многомасштабной вычислительной схемы. Можно выделить три масштабных уровня. На первом уровне осуществляется квантово-механическое моделирование, в результате которого определяется энергия активации кислородных вакансий, а также величины, которые используются для построения межатомного потенциала. На втором уровне с помощью построенного потенциала осуществляется молекулярно-динамическое моделирование, целью которого является определение энергетических барьеров миграции ионов кислорода. Методы расчета на первых двух уровнях описаны в разделе 1.2. На третьем уровне (раздел 1.3) выполняется моделирование динамики ионов методом Монте-Карло с использованием полученных миграционных барьеров. В разделе 1.4 эта многомасштабная схема моделирования применяется к оксиду кремния. Исследуемая структура представляет собой матрицу альфа-кварца между парой инертных электродов. Представлены результаты вычислительных экспериментов, демонстрирующих стадии роста и разрушения проводящих филаментов под действием переменного во времени напряжения.

Модели второго класса – приближенные (полуэмпирические). Они основаны на представлении оксидного слоя как нелинейной динамической системы относительно переменной состояния, связанной с концентрацией кислородных вакансий [21–27]. Параметры модели определяются путем минимизации отклонения расчетных вольт-амперных характеристик от наблюдаемых в экспериментах или полученных на основе прямого

моделирования. Модели данного класса описаны в главе 2. Такие модели, будучи достаточно экономичными в вычислительном отношении, могут быть эффективно использованы при имитационном моделировании нейроморфных систем на основе мемристорных кросс-баров [28, 29]. Этот подход представлен в главе 3.

Присущее мемристорным кроссбарам свойство естественной параллелизации матрично-векторных операций создает возможности для их эффективного использования в аналоговой реализации классических нейронных сетей. Аналоговые вычисления производятся на порядки быстрее по сравнению с вычислениями на центральном процессоре и на графических ускорителях. Кроме того, существенно ниже энергозатраты на проведение математических операций. При этом существенной особенностью аналоговых вычислений является небольшая точность. В главе 4 рассматривается аналоговая реализация классических нейронных сетей, исследуется зависимость качества работы классических нейронных сетей от точности задания весов.

В главе 5 показана применимость интеграционной платформы для проведения многомасштабных расчетов нейроморфных сетей от атомного уровня до макроуровня согласно специально подготовленным сценариям, что позволяет автоматизировать процесс оптимизации их работы и уменьшить временные затраты на получение новых результатов.

ГЛАВА 1

МОДЕЛИРОВАНИЕ ОБРАЗОВАНИЯ/РАЗРУШЕНИЯ ПРОВОДЯЩИХ ФИЛАМЕНТОВ В МЕМРИСТИВНЫХ ЭЛЕМЕНТАХ

1.1. Многомасштабная вычислительная модель

Расчетная схема многомасштабного моделирования динамики филаментных структур в наноразмерных оксидных слоях представлена на **рисунке 1.1**.

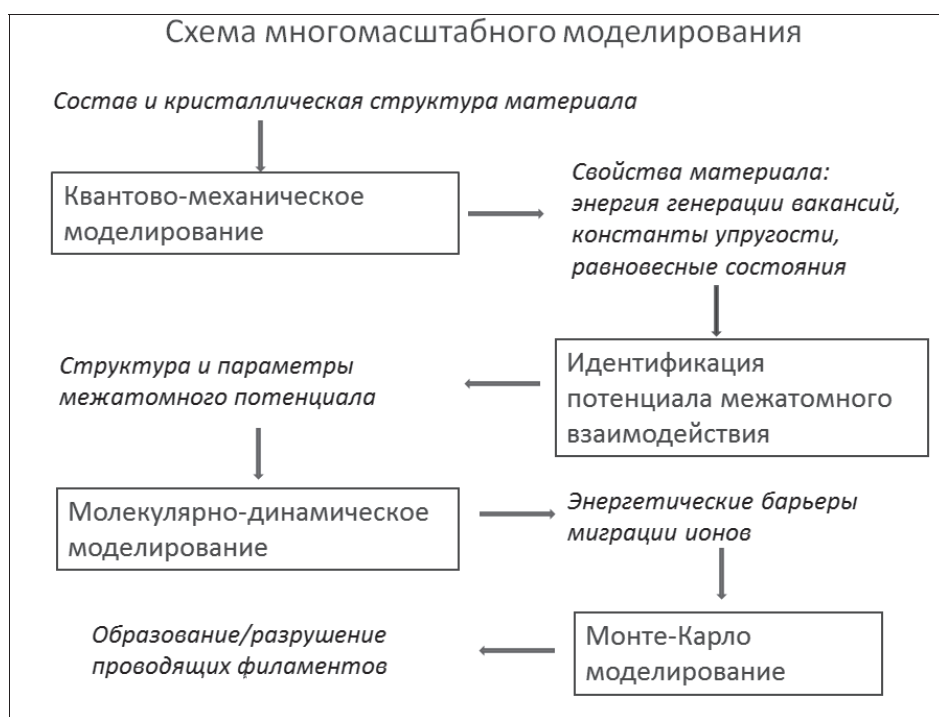


Рис. 1.1. Вычислительная схема

Можно выделить следующие основные этапы многомасштабного моделирования. На первом этапе проводятся квантово-механические (первопринципные) расчеты. Это позволяет определить распределение электронной плотности в структуре и энергетические барьеры для генерации/рекомбинации ионов кислорода. По результатам

первопринципного моделирования генерируется информация для идентификации эмпирических межатомных потенциалов. Они используются на втором этапе, где проводится молекулярно-динамическое моделирование. Для параметрической идентификации межатомных потенциалов в настоящей работе реализованы алгоритмы градиентной оптимизации с применением быстрого автоматического дифференцирования. Использование молекулярно-динамического моделирования непосредственно в данной задаче затруднено из-за несовместимости характерных временных масштабов. Однако моделирование молекулярной динамики может быть использовано для расчета энергетических барьеров миграции ионов кислорода. Здесь нами реализован подход, сочетающий метод эластичной ленты с методом сопряженных градиентов для релаксации атомарной структуры. Информация об энергетических барьерах передается на следующий уровень, где выполняется моделирование методом Монте-Карло динамики ионов в мемристивном элементе. На данном этапе проводится статистическое моделирование на основе случайной реализации следующих событий: генерация свободного иона путем разрыва связи с кристаллической решеткой, рекомбинация свободного иона и незанятой вакансии, миграция ионов по узлам и междоузлиям кристаллической решетки. Вероятности событий рассчитываются на основе энергетических барьеров, определяемых с помощью молекулярно-динамического моделирования. Для согласования электрического поля с распределением ионов и вакансий в мемристивном элементе на каждом шаге кинетического метода Монте-Карло решается трехмерное уравнение Пуассона.

1.2. Квантово-механическое и молекулярно-динамическое моделирование

Расчеты из первых принципов являются наиболее точным доступным способом моделирования атомных систем. Единственная необходимая информация для расчетов из первых принципов – это структура кристаллической решетки. Зная приблизительную структуру решетки, можно

уточнить ее и рассчитать различные свойства, такие как энергия когезии, упругие свойства, силы, действующие на атомы в несовершенной (при наличии дефектов) решетке. Одним из важнейших параметров, подлежащих расчету на этом этапе, является энергия образования вакансии и иона. Для образования заряженной пары вакансия-ион энергия может быть рассчитана следующим образом [31, 32]:

$$E_g = E_v - E_b + \mu + qE_f$$

где E_b – энергия идеальной кристаллической системы, E_v – энергия системы с вакансией, μ – химический потенциал движущегося атома, q – зарядовое состояние вакансии, E_f – энергия Ферми.

В настоящей работе для проведения первопринципных расчетов используется программное обеспечение с открытым исходным кодом Quantum Espresso.

Расчеты из первых принципов точны, но при этом чрезвычайно затратны в вычислительном отношении. Моделирование больших масштабов требует использования различных приближений. Одним из мощных средств расширения масштаба является молекулярно-динамическое моделирование. Для его проведения необходимо найти соответствующий межатомный потенциал. Эмпирические потенциалы могут иметь различные формы: от довольно простых, таких как парный потенциал Леннарда-Джонса, который широко используется для моделирования простых взаимодействий, до различных сложных многочастичных потенциалов, таких как вариации потенциала Терсоффа или модели встроенного атома, специализированных силовых полей для конкретных материалов. Структура потенциала всегда выбирается исходя из типа взаимодействий в моделируемой системе. После выбора структуры необходимо определить параметры потенциала. Идентификация параметров представляет собой глобальную оптимизационную задачу, как правило, со сложным рельефом поверхности целевой функции. Первоначально проводится грубая глобальная оптимизация, которая может быть выполнена с использованием

метавристических методов, таких как, например, имитация отжига, после чего осуществляется локальная оптимизация. Для локальной оптимизации в настоящей работе используется метод BFGS (Бройден, Флетчер, Гольдфарб, Шанно), который требует информации о градиентах функции. Градиенты функции извлекаются с использованием метода автоматического дифференцирования, поскольку дифференцировать потенциал вручную затруднительно, а численное дифференцирование с использованием конечных разностей вычислительно не эффективно. Чтобы использовать процедуры оптимизации, все важные свойства, извлеченные из первопринципных расчетов, объединяются во взвешенную целевую квадратичную функцию, которая минимизируется:

$$J(\theta) = \sum_j \left(\frac{P_j(\theta) - P_j^*}{w_j} \right)^2$$

$$\theta^* = \arg \min_{\theta} J(\theta)$$

Здесь J – целевая функция, j – номер учитываемого свойства, θ – набор параметров потенциала, P_j – значение j -го свойства, вычисляемое по молекулярно-динамической модели с соответствующим набором параметров, P_j^* – эталонное значение j -го свойства, определенное из первопринципных расчетов, w_j – вес j -го свойства. Набор свойств включает энергию когезии, упругие свойства и силы, действующие на атомы в неидеальной решетке.

Как уже упоминалось во введении, прямое использование молекулярно-динамического моделирования в этой задаче затруднено из-за несовместимости характерных временных масштабов. Частоты генерации вакансий и актов миграции ионов на несколько порядков ниже частоты колебаний атомов. Однако, возможно использовать молекулярно-динамическое моделирование по-другому. Имея потенциал межатомного взаимодействия с подобранными параметрами, мы можем использовать его для определения барьеров миграции, необходимых для кинетического метода Монте-Карло. Для определения миграционного барьера необходимо найти

седловую точку энергетической поверхности, расположенную между двумя стабильными позициями ионов. Это можно сделать с помощью разных методов.

Простейшим методом для определения пути с минимальной энергией является метод принудительного волочения атома. Для вычисления энергетического барьера методом волочения выбирается начальное и конечное положение некоторого атома. Задается число шагов метода, и атом постепенно, шаг за шагом сдвигается из начальной точки в конечную. На каждом шаге вычисляется направление от текущего положения исследуемого атома на конечную точку его пути ($\mathbf{d}_i = \mathbf{r}_i^* - \mathbf{r}_i$), оставшееся количество шагов (k), и атом сдвигается на \mathbf{d}_i/k . Затем производится релаксация системы с дополнительным ограничением, состоящим в том, что исследуемый атом может двигаться только в плоскости ортогональной вектору, на который он был сдвинут (\mathbf{d}_i). Метод является довольно простым в реализации, а также не требует слишком больших вычислительных затрат. Однако в нём важно следить за общей динамикой системы, так как перемещение одного из атомов может приводить к перемещению соседей, а в наиболее сложных случаях к комплексному изменению структуры системы.

Для решения проблемы с существенным изменением структуры системы в качестве альтернативы можно использовать метод эластичной ленты [32]. Данный метод позволяет уменьшить величину случайных перемещений атомов. Вместо постоянного перемещения атома в направлении новой позиции проводится совместное связанное молекулярно-динамическое моделирование множества копий исследуемой системы. На первом этапе проводится релаксация двух копий: копии системы с атомами на своих начальных позициях и копии системы с атомами на своих конечных позициях. Затем позиции атомов интерполируются между начальным и конечным состоянием для создания множества промежуточных конфигураций. Все полученные конфигурации объединяются в одну общую систему таким образом, что все атомы из одной конфигурации взаимодействуют между собой с помощью исследуемого потенциала, а также

соответствующие атомы из каждой конфигурации объединяются в одну цепочку, объединённую виртуальными пружинами. В результате силы, действующие на атомы, определяются не только их пространственным соседством, но и их временными соседями (**рисунок 1.2**). При этом силы, действующие на атом внутри системы, проецируются на плоскость перпендикулярную виртуальной цепочке, а на силы, действующие на атомы со стороны пружины накладывается существенный штраф, ограничивающий перемещение атомов в этой плоскости. Таким образом, атомы оказываются более ограничены в своих перемещениях, что позволяет более надёжно определять пути минимальной энергии.

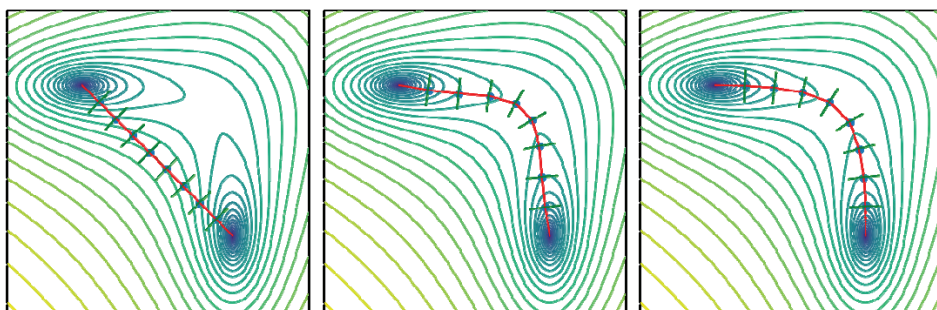
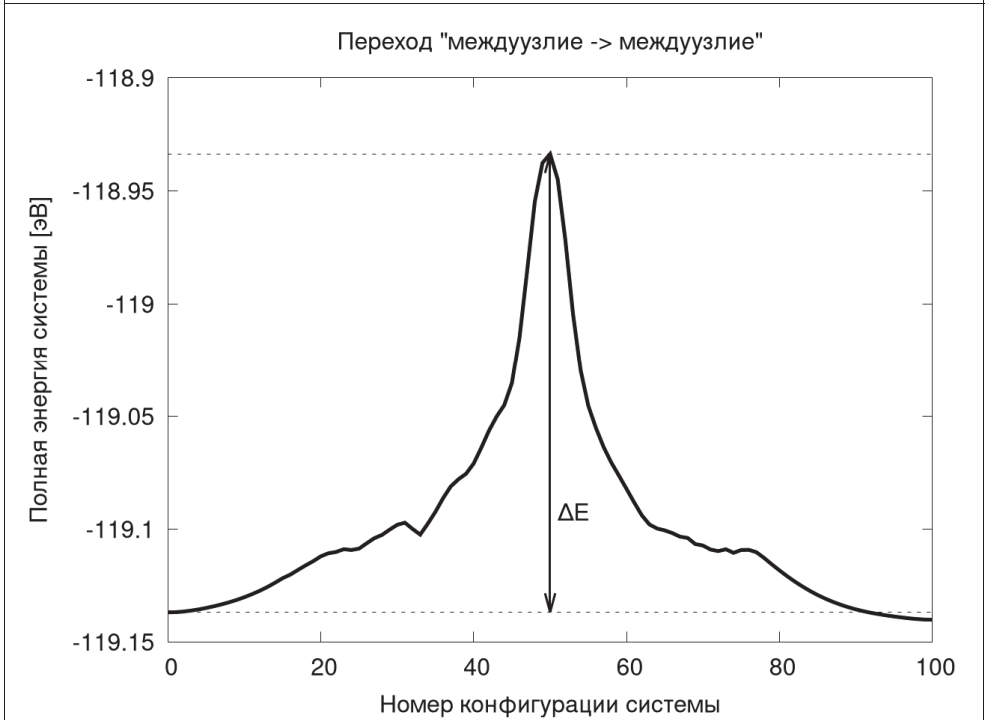
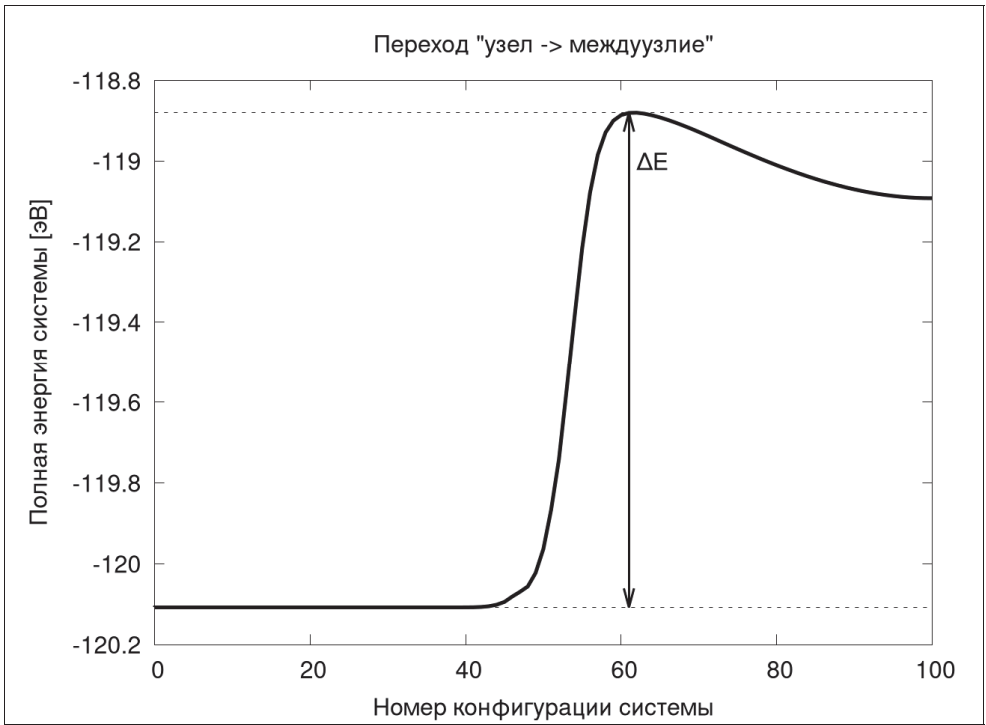


Рис. 1.2. Визуализация метода эластичной ленты на искусственной энергетической поверхности

Используя данный метод, можно определить энергетические барьеры переходов атомов из узлов кристаллической решётки в междуузлия, между междуузлиями и из междуузлий обратно в узлы кристаллической решётки, если известны позиции междуузлий.

Рассмотрим модельный пример для материала, имеющего гранецентрированную решётку, атомы которого связаны потенциалом Морзе. Для гранецентрированной решётки в каждой элементарной ячейке геометрически можно определить 4 октаэдрических и 8 тетраэдрических устойчивых междуузлий. На **рисунке 1.3** показаны определенные методом эластичной ленты энергетические барьеры переходов «узел-междузлье», «междузлье-междузлье», «междузлье-узел».



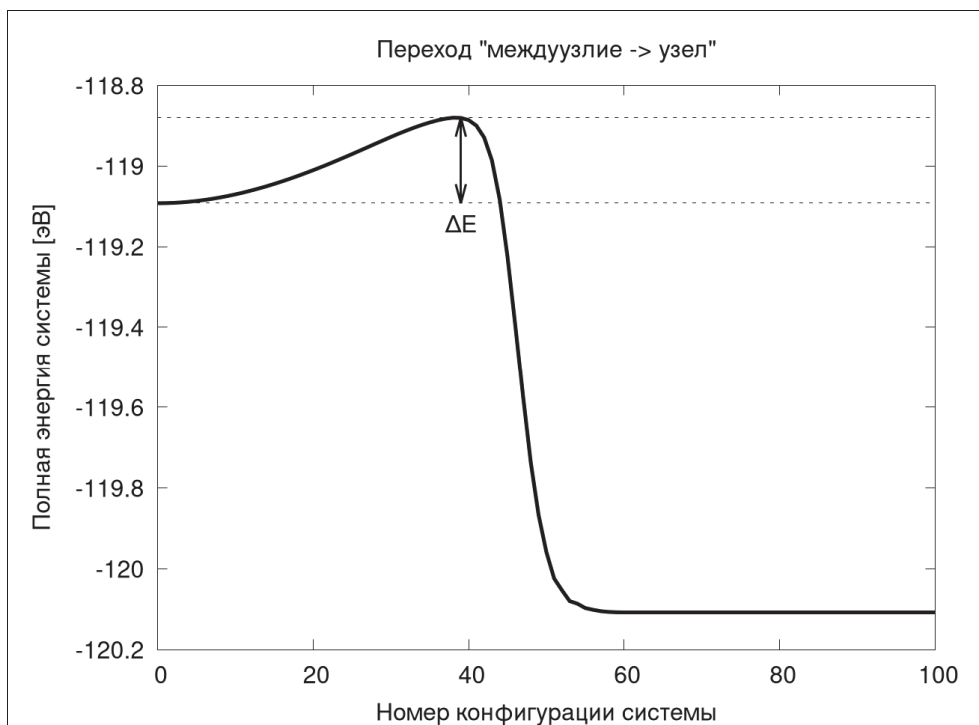


Рис. 1.3. Энергетические барьеры

Существенным элементом молекулярно-динамических методов определения энергетических барьеров является релаксация атомарной структуры. Аналитически предсказать устойчивое положение для системы атомов, связанных нетривиальными потенциалами, в общем случае невозможно. Для того чтобы получить состояние, соответствующее минимуму энергии, можно применять алгоритмы основанные как на физической интерпретации задачи, так и алгоритмы оптимизации общего вида.

Наиболее очевидным методом является решение уравнений динамики до достижения состояния равновесия (метод quickmin). В ходе вычислительного процесса проводится обычное молекулярно-динамическое моделирование системы с единственным ограничением, что если в какой-то момент времени скорость атома и сила, действующая на него, направлены в

противоположные стороны, то скорость атома зануляется. Таким образом, атомы быстро приближаются к локальным минимумам, а затем резко тормозят, как только начинают из него выходить. Таким образом, решается следующая задача:

$$\begin{aligned}\dot{\mathbf{r}}_i &= \mathbf{v}_i \\ m_i \dot{\mathbf{v}}_i &= \mathbf{F}_i \\ \mathbf{F}_i &= -\nabla_i G\end{aligned}$$

где m_i – масса i -го атома, \mathbf{r}_i – радиус-вектор, \mathbf{v}_i – вектор скорости, \mathbf{F}_i – сила, действующая на i -й атом, G – межатомный потенциал. Задача решается методом скоростей Верле:

$$\begin{cases} \mathbf{r}_i^{k+1} = \mathbf{r}_i^k + \Delta t \mathbf{v}_i^k + \frac{\Delta t^2}{2m_i} \mathbf{F}_i^k \\ \mathbf{F}_i^{k+1} = -\nabla_i G^{k+1} \\ \mathbf{v}_i^{k+1} = \mathbf{v}_i^k + \frac{\Delta t}{2m_i} (\mathbf{F}_i^k + \mathbf{F}_i^{k+1}) \end{cases}$$

где Δt – шаг по времени.

Для стабилизации решения используются методы демпфирования, такие как постоянный отток энергии $\mathbf{v}_i^{*k+1} = \alpha \mathbf{v}_i^{k+1}$ или полная остановка атома, если он попытается подняться на энергетическую поверхность:

$$\mathbf{v}_i^{*k+1} = \begin{cases} \mathbf{v}_i^{k+1}, & (\mathbf{v}_i^{k+1}, \mathbf{F}_i^{k+1}) > 0 \\ 0, & (\mathbf{v}_i^{k+1}, \mathbf{F}_i^{k+1}) \leq 0 \end{cases}$$

Минусом этого метода является необходимость задания всех необходимых для молекулярной динамики параметров, которые включают в себя массы частиц и шаг интегрирования по времени. От последнего может существенно зависеть скорость сходимости метода, а также проявление неустойчивости при далеких от равновесия начальных условиях.

В качестве альтернативы методу quickmin возможно использование методов, определяющих позиции атомов, которые обеспечивают минимум потенциальной энергии системы. Из стандартных методов оптимизации хорошо зарекомендовал себя метод сопряжённых градиентов, который не

требует большого количества дополнительной памяти и вычислительных ресурсов, а также не требует дополнительных данных, таких как массы атомов и временные параметры.

Сравним методы `quickmin` и сопряженных градиентов на двух конкретных задачах, чтобы получить представление о реальной производительности. В качестве первой задачи возьмём задачу релаксации поверхности. Пусть имеется идеальный кристалл, периодический в двух направлениях и имеющий свободную поверхность в третьем направлении. Равновесное расстояние между слоями у поверхности не будет равно равновесному расстоянию между слоями внутри идеального кристалла, поэтому после расстановки в формации идеального кристалла эта система несколько деформируется в направлении, перпендикулярном свободным поверхностям. На **рисунке 1.4** изображён процесс сходимости двух методов для этой системы.

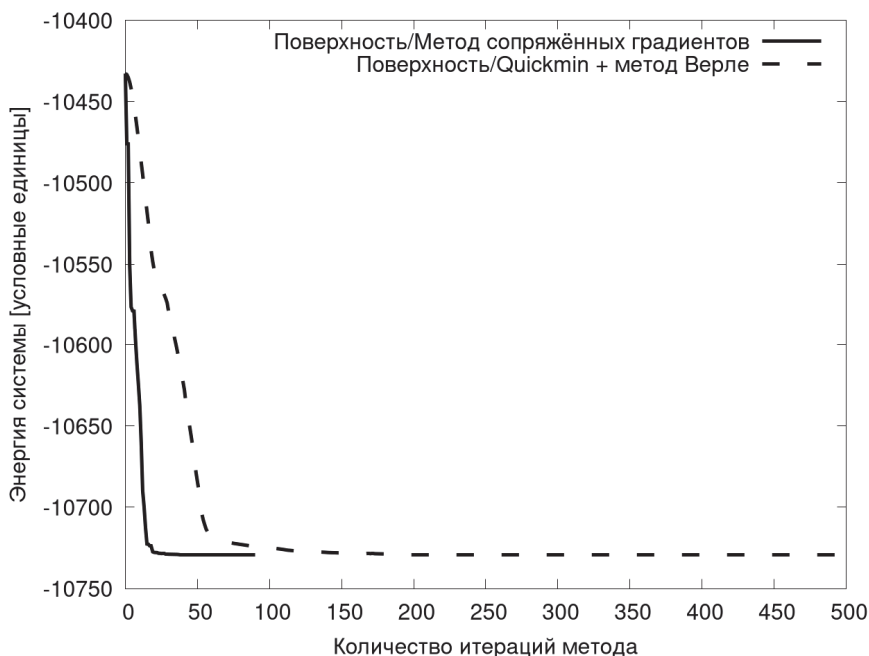


Рис. 1.4. Сходимость методов релаксации поверхностной атомарной структуры

В качестве второй задачи рассмотрим задачу релаксации вакансии в кристалле. Пусть имеется идеальный периодический кристалл, из которого был извлечён один из атомов. Атомы, соседствующие с извлечённым, будут вынуждены некоторым образом переместиться, чтобы вновь оказаться в локальном минимуме энергии. На **рисунке 1.5** изображён процесс сходимости двух методов для этой системы.

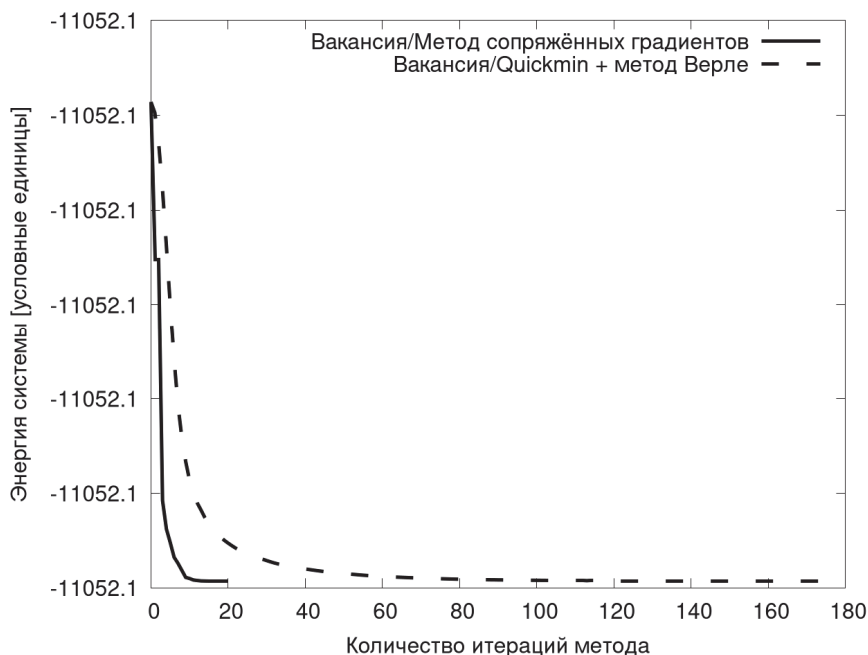


Рис. 1.5. Сходимость методов релаксации объемной атомарной структуры

Видно, что метод сопряжённых градиентов сходится существенно быстрее quickmin. Однако не всё определяется только количеством итераций методов. Сложность итераций метода сопряжённых градиентов существенно выше, так как внутри производится линейный поиск вдоль выбранного направления оптимизации. Таким образом, отдельная итерация метода сопряжённых градиентов включает в себя большее количество вычислений, чем отдельная итерация quickmin. На **рисунке 1.6** изображено количество вызовов функции расчёта энергии методом сопряжённых градиентов.

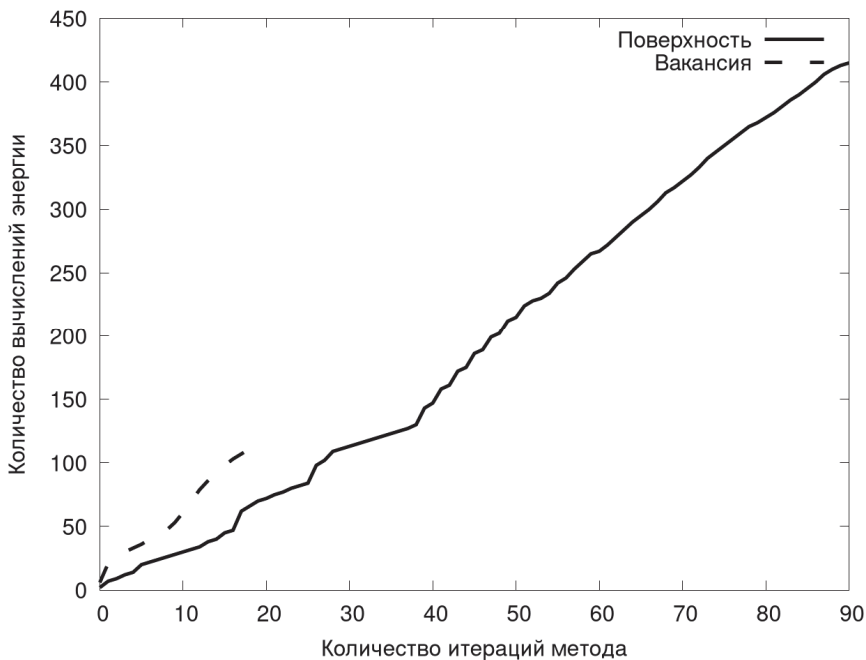


Рис. 1.6. Число вызовов функции вычисления энергии в методе сопряженных градиентов

Можно заметить, что количество вычислений энергии методом сопряжённых градиентов близко к количеству итераций метода quickmin. Однако в данных задачах метод сопряжённых градиентов работает в 2–10 раз быстрее, чем метод quickmin, так как, несмотря на близкое количество вычислений функции энергии, quickmin всякий раз вычисляет силы, действующие на атомы, что бывает заметно дороже, чем простое вычисление энергии систем.

1.3. Монте-Карло моделирование ионной динамики в мемристивных элементах

Метод молекулярной динамики, будучи значительно дешевле расчетов из первых принципов, тем не менее, требует огромных вычислительных ресурсов, когда необходимо исследовать поведение системы с большим количеством редких событий. Например, процесс образования ионов кислорода в оксидной матрице под действием приложенного напряжения

может занять несколько секунд, чтобы образовался один ион, и такие временные масштабы не доступны в рамках молекулярно-динамического моделирования. Кинетический метод Монте-Карло позволяет исследовать поведение различных систем на относительно больших временных масштабах. Вместо того, чтобы моделировать каждое движение каждого атома в системе, мы определяем интенсивность каждого интересующего события и моделируем не движения частиц, а поток событий. События могут происходить довольно редко, но каждое событие продвигает модель вперед на определенное время. Для мемристивных структур необходимо учитывать три события, связанные с ионами:

1. Генерация пары ион-вакансия
2. Миграция ионов между двумя устойчивыми позициями (узлы, междоузлия решетки).
3. Рекомбинация пар ион-вакансия

Каждое событие имеет связанный с ним энергетический барьер, который ион должен преодолеть, чтобы произошел переход.

Интенсивность событий генерации может быть описана следующим соотношением (см., напр., [18]):

$$R_g = f_0 \exp\left(-\frac{E_g - \gamma eE}{kT}\right)$$

где R_g – интенсивность генерации, f_0 – частота колебаний атомов (частота Дебая), E_g – барьер генерации, γ – вклад поляризации связи в локальное электрическое поле, e – заряд электрона, E – модуль напряженности электрического поля, k – постоянная Больцмана, T – температура.

Интенсивность событий рекомбинации можно описать с помощью:

$$R_r = f_0 \exp\left(-\frac{E_r - \gamma eE}{kT}\right)$$

где R_r – скорость рекомбинации, E_r – барьер рекомбинации.

Интенсивность событий миграции (дрейфа-диффузии) можно описать следующим образом:

$$R_m = f_0 \exp\left(-\frac{E_m - \Delta E}{kT}\right)$$

где R_m – скорость миграции, E_m – барьер миграции, ΔE – разница между энергиями стартового и конечного положения атомов.

На каждом шаге кинетического метода Монте-Карло система переходит в один из соседних локальных минимумов энергии в соответствии с распределением вероятностей перехода. Таким образом, на каждом шаге моделирования изменяется состояние системы, что позволяет существенно ускорить моделирование по сравнению с методом молекулярной динамики. Для выбора следующего состояния заметим, что время до свершения событий подчиняется экспоненциальному распределению. Данный факт позволяет воспользоваться следующими свойствами данного распределения. Пусть система находится в состоянии i и известны все частоты переходов в соседние состояния j равные λ_{ij} , тогда:

1. тогда времена переходов распределены по экспоненциальному закону $\sim \exp(-\lambda_{ij})$;
2. время самого раннего события распределено по закону $\exp\left(-\sum_j \lambda_{ij}\right)$;
3. вероятность того, что событие j произойдёт раньше остальных равна $\lambda_{ij} \left(\sum_j \lambda_{ij}\right)^{-1}$.

Для определения времени ближайшего события достаточно сгенерировать экспоненциально распределённую случайную величину. Для этого можно воспользоваться методом обратной функции, сгенерировав равномерно распределённую случайную величину и подставив её в функцию обратной функции распределения экспоненциальной величины:

$$x \sim U(0,1)$$

$$-\lambda^{-1} \ln(x) \sim \text{Exp}(\lambda)$$

Для выбора же самого события также можно применить метод обратной функции, однако в данном случае функция распределения будет

кусочно-постоянной, и выбор следующего состояния j должен происходить исходя из следующего соотношения:

$$x \sim U(0,1)$$

$$\sum_{k=1}^j \lambda_{ik} \leq x \sum_{k=1}^n \lambda_{ik} \leq \sum_{k=1}^{j+1} \lambda_{ik}$$

В качестве наиболее полного варианта статистического моделирования нами реализована организация перемещения ионов кислорода по узлам и междоузлиям реальной кристаллической решетки. Разумеется, применение такого подхода связано с высокими вычислительными затратами. Поэтому для проведения предварительного анализа были реализованы облегченные варианты моделирования ионного транспорта в оксидном слое.

В первом варианте по-прежнему осуществляется полномасштабное Монте-Карло моделирование миграции всех ионов, находящихся в системе со случайным выбором времени наступления события, однако движение осуществляется не по реальной кристаллической решетке, а путем сдвига позиции иона на постоянную решетки вдоль одной из координатных осей, случайно выбранной в соответствии с вероятностями переходов. Поскольку в данном варианте осуществляется движение ионов вдоль координатных осей, такой подход может быть охарактеризован как покоординатное случайное блуждание.

Другой, более экономичный, вариант заключается в численном решении макроскопического уравнения дрейфа-диффузии:

$$\frac{\partial C(\mathbf{r}, t)}{\partial t} = \nabla \cdot (D(T) \nabla C(\mathbf{r}, t) - \mathbf{V}(\mathbf{E}, T) C(\mathbf{r}, t)) + g(\mathbf{r}, t)$$

При этом выражения, определяющие осредненные вектор скорости и коэффициент диффузии в зависимости от температуры и напряженности электрического поля, связаны с интенсивностями соответствующих событий:

$$\mathbf{V}(\mathbf{E}, T) = af_0 \exp\left(-\frac{E_m}{k_B T(\mathbf{r})}\right) \sinh\left(\frac{eaE(\mathbf{r})}{2k_B T(\mathbf{r})}\right) \frac{\mathbf{E}(\mathbf{r})}{E(\mathbf{r})}$$

$$D(T) = \frac{1}{2} a^2 f_0 \exp\left(-\frac{E_m}{k_B T(\mathbf{r})}\right)$$

Здесь a – постоянная решетки, $g(\mathbf{r})$ – распределение плотности генерации ионов кислорода по объему, $C(\mathbf{r})$ – распределение концентрации ионов кислорода по объему оксидного слоя.

Эффективным методом решения уравнения дрейфа-диффузии в данных условиях является Монте-Карло моделирование динамики вычислительных частиц со случайным выбором направления движения в соответствии с вероятностями, определяемыми коэффициентами уравнения (случайное блуждание по вычислительной сетке). Будучи микроскопическим по форме, этот метод остается макроскопическим по существу, а точность итогового решения зависит только от выбранного числа симуляций. Такой подход позволяет задавать различное поведение ионов на границах области, включая электроды, оставаясь в рамках приемлемых вычислительных затрат.

Образование кислородных вакансий создает условия для туннелирования электронов и протекания электрического тока в оксидном слое. В общем случае моделирование этого эффекта осуществляется методом Монте-Карло с учетом различных механизмов туннелирования электронов. Интерес представляют три события, связанных с динамикой электронов:

1. Инжекция электронов с электрода в ловушку;
2. Туннелирование электронов между ловушек;
3. Переход электронов из ловушек на электрод.

Эти переходы можно описать аналогично ионным переходам. Скорость туннелирования от ловушки к ловушке с энергиями E_1 и E_2 , когда E_1 больше E_2 , может быть описана следующим соотношением [18].

$$R_t = f_0 \exp\left(-\frac{2R}{\xi}\right)$$

где R – расстояние между ловушками, ξ – длина локализации электрона. Этот переход вызовет испускание фононов с полной энергией $E_1 - E_2$. Когда E_1

меньше E_2 , происходит поглощение фононов, и скорость туннелирования можно описать следующим образом:

$$R_t = f_0 \exp\left(-\frac{2R}{\xi}\right) \exp\left(\frac{E_1 - E_2}{\xi}\right)$$

Как и для ионной динамики, здесь на этапе предварительного анализа можно использовать более простую модель, учитывающую зависимость локальной электрической проводимости от концентрации вакансий кислорода

$$\mathbf{J}(\mathbf{r}) = \sigma(C_{vac})\mathbf{E}(\mathbf{r})$$

$C_{vac}(\mathbf{r})$ – распределение концентрации кислородных вакансий по объему

Для функциональной зависимости проводимости от концентрации вакансий можно использовать следующее выражение:

$$\sigma(C_{vac}) = \frac{\sigma_h}{1 + \frac{\sigma_h - \sigma_l}{\sigma_l} \exp(-\alpha\sigma_h C_{vac})}$$

где σ_l , σ_h – низкая и высокая проводимости материала, параметр α характеризует резкость перехода от состояния с низкой проводимостью к состоянию с высокой проводимостью.

Динамика заряженных частиц создает изменения электрического поля внутри моделируемого объема. Этот эффект необходимо учитывать для получения самосогласованного решения. Для расчета электрического поля решается трехмерное уравнение Пуассона. Следует отметить, что в наиболее полном варианте моделирования необходимо решать уравнение Пуассона после каждого события, связанного с миграцией ионов или электронов, что является достаточно дорогостоящей процедурой. Чтобы снизить вычислительную сложность, нами разработан специальный алгоритм. Уравнение Пуассона для распределения плотности заряда в прямоугольном объеме между двумя пластинами можно записать следующим образом:

$$\begin{cases} \Delta\phi = -\frac{\rho}{\varepsilon} \\ \phi(0, y, z) = \phi(l_x, y, z) \\ \phi(x, 0, z) = \phi(x, l_y, z) \\ \phi(x, y, 0) = 0 \\ \phi(x, y, l_z) = U \end{cases}$$

Здесь ϕ – электростатический потенциал, ρ – плотность заряда, ε – диэлектрическая проницаемость, l_x, l_y , – размеры объема в направлениях x, y (параллельно электродам) и z (нормаль к электродам), U – внешнее напряжение. Далее воспользуемся тем, что задача периодична по направлениям x, y . Решение задачи можно представить в виде суперпозиции двух решений: решения задачи с нулевыми граничными условиями и решения задачи с нулевой плотностью внутреннего заряда:

$$\begin{cases} \Delta\phi = 0 \\ \phi(0, y, z) = \phi(l_x, y, z) \\ \phi(x, 0, z) = \phi(x, l_y, z) \\ \phi(x, y, 0) = 0 \\ \phi(x, y, l_z) = U \end{cases} \begin{cases} \Delta\phi = -\frac{\rho}{\varepsilon} \\ \phi(0, y, z) = \phi(l_x, y, z) \\ \phi(x, 0, z) = \phi(x, l_y, z) \\ \phi(x, y, 0) = 0 \\ \phi(x, y, l_z) = 0 \end{cases}$$

Первую задачу легко решить, имеет место линейное решение:

$$\phi(x, y, z) = U \frac{z}{l_z}. \quad \text{Для решения второй задачи учтем периодичность.}$$

Конфигурация объема моделирования и периодические граничные условия позволяют использовать подход, основанный на функциях Грина. В качестве этапа предварительной обработки вычисляется набор функций Грина для зарядов в одном столбце расчетной сетки. Функции Грина получены спектральным методом с использованием быстрого преобразования Фурье:

$$\begin{aligned} B_{k_x k_y k_z} &= \exp\left(-2\pi i \frac{x}{l_x} k_x\right) \exp\left(-2\pi i \frac{y}{l_y} k_y\right) \sin\left(\pi \frac{z}{l_z} k_z\right) \\ \sum_{\substack{k_x, k_y \in Z \\ k_z \in N}} -a_{k_x k_y k_z} \left(\left(\frac{2\pi k_x}{l_x}\right)^2 + \left(\frac{2\pi k_y}{l_y}\right)^2 + \left(\frac{2\pi k_z}{l_z}\right)^2 \right) B_{k_x k_y k_z}(x, y, z) &= \sum_{\substack{k_x, k_y \in Z \\ k_z \in N}} b_{k_x k_y k_z} B_{k_x k_y k_z}(x, y, z) \\ a_{k_x k_y k_z} &= -b_{k_x k_y k_z} \left(\left(\frac{2\pi k_x}{l_x}\right)^2 + \left(\frac{2\pi k_y}{l_y}\right)^2 + \left(\frac{2\pi k_z}{l_z}\right)^2 \right)^{-1} \end{aligned}$$

Решения для зарядов в разных столбцах будут трансляционно-инвариантными из-за периодической структуры задачи. Каждый раз, когда заряженная частица пересекает границу между ячейками в расчетной сетке, искомые функции Грина находятся в наборе предварительно вычисленных решений, масштабированных в соответствии с зарядом частицы, и решение уравнения Пуассона модифицируется простым сложением и вычитанием масштабированных функций Грина. Следует отметить, что такая процедура требует только линейного количества операций по отношению к количеству ячеек.

Вообще говоря, вместе с уравнением Пуассона необходимо решать уравнение теплопередачи, чтобы учесть температурное поле, влияющее на скорость указанных выше событий. Температурное поле определяется Джоулевым теплом и теплоотводом на границах. Уравнение теплопроводности имеет вид:

$$c\rho \frac{\partial T(\mathbf{r}, t)}{\partial t} = \nabla \cdot (\lambda(T)\nabla T(\mathbf{r}, t)) + \mathbf{J}(\mathbf{r})\mathbf{E}(\mathbf{r})$$

$T(\mathbf{r})$ – распределение температуры по объему (температурное поле передается в, c - удельная теплоемкость материала, ρ – плотность материала, λ – коэффициент теплопроводности материала, $\mathbf{J}(\mathbf{r})$ – распределение плотности потока электронов по объему, $\mathbf{E}(\mathbf{r})$ – распределение вектора напряженности электрического поля по объему слоя.

Оценки показывают, что характерное время тепловой релаксации существенно ниже характерных времен генерации, дрейфа, диффузии ионов, что позволяет использовать стационарное уравнение теплопереноса

$$\nabla \cdot (\lambda(T)\nabla T(\mathbf{r}, t)) + \mathbf{J}(\mathbf{r})\mathbf{E}(\mathbf{r}) = 0$$

Уравнение дополняется граничными условиями, задающими интенсивность теплоотвода с границ рассматриваемого объема.

1.4. Вычислительные эксперименты

Вычислительные эксперименты проводились для мемристового элемента, в котором в качестве активной среды используется оксид кремния SiO₂.

Сначала рассмотрим результаты двумерного моделирования с использованием сеточных вычислительных моделей дрейфа-диффузии ионов и электрического тока. При таком подходе учитывается характерный размер кристаллической решетки, однако реальная структура решетки не учитывается. Размеры исследуемой области приняты равными 20 нм (ширина) и 10 нм (высота), число ячеек разбиения в соответствующих направлениях 21x11, число атомарных позиций кислорода в вычислительной ячейке – 8. В качестве начальной концентрации кислородных вакансий задаются относительные величины для объема в целом и для выделенной (дефектной) области, вытянутой по высоте мемристора. Для применения сеточной модели электрического тока должны быть заданы проводимости оксидного слоя, соответствующие состояниям высокой и низкой резистивности. Они приняты равными 100 и 1000 См/м соответственно. Для расчета температурного поля в элементе необходимо задание граничных условий. Без знания схемы охлаждения элемента это затруднительно. Здесь для определенности на каждой границе заданы коэффициент теплоотдачи 1000 Втм⁻²К⁻¹ и температура окружающей среды 300 К.

Остальные величины, используемые в расчетах, следующие: коэффициент теплопроводности оксида принят равным 5.2 Втм⁻¹К⁻¹, относительный коэффициент диэлектрической проницаемости – 10.2, частота Дебая – 10¹³ с⁻¹, постоянная решетки – 0.5 нм, коэффициент, учитывающий вклад поляризации связи в энергетический барьер – 7*10⁻¹⁰ эВ, энергия активации процесса миграции ионов кислорода – 0.3 эВ, энергия активации процессов генерации и рекомбинации ионов кислорода – 1 эВ. Временной интервал моделирования – 1 мс, электростатические потенциалы на нижнем и верхнем электродах – 3.2, 0 В. На электродах задано условие полного отражения ионов кислорода.

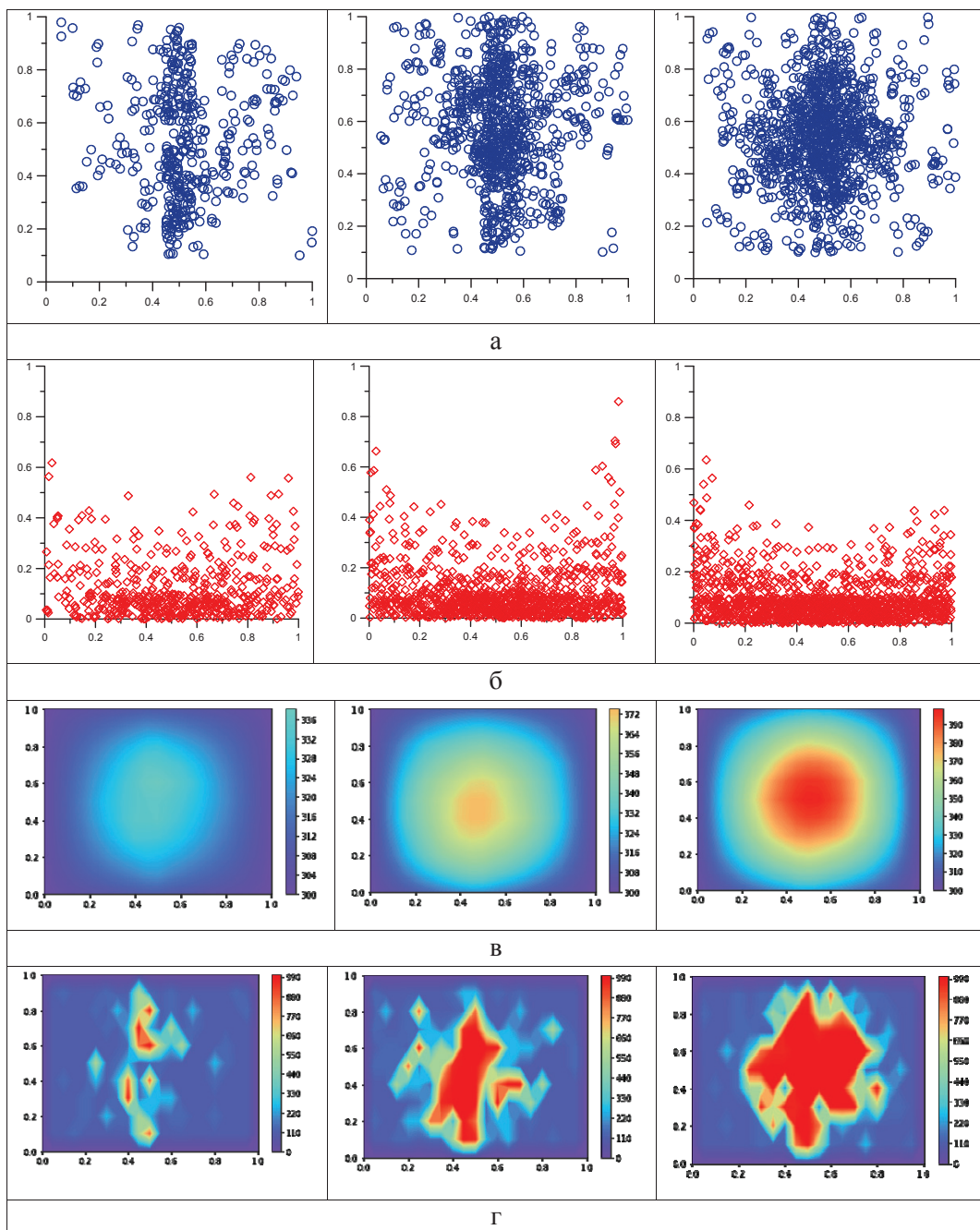


Рис. 1.7. Временная развертка процесса образования филамента: а – кислородные вакансии, б – ионы кислорода, в – температура, г – проводимость

На **рисунке 1.7** представлена типичная временная развертка процесса для случая задания выделенной (дефектной) области, вытянутой по высоте оксидного слоя, соответствующей межкристаллитному пространству. Представлены распределение кислородных вакансий (синий цвет), распределение ионов кислорода (красный цвет), поле температур и распределение локальной проводимости. По осям отложены относительные координаты.

Относительно перехода в низкоомное состояние наблюдается следующая картина: подается управляющее воздействие в виде разности потенциалов на верхнем и нижнем электродах, что ослабляет активационные барьеры генерации кислородных вакансий и интенсифицирует процесс дрейфа ионов. В результате ионы дрейфуют к нижнему электроду, что существенно снижает интенсивность их рекомбинации в большей части объема, тогда как процесс генерации вакансий протекает без помех. Эффект усиливается за счет положительной обратной связи, повышение содержания вакансий приводит к усилению тока, что в свою очередь ведет к разогреву, и, как следствие, интенсификации процессов генерации вакансий и дрейфа-диффузии ионов. Снятие напряжения фиксирует сформировавшуюся структуру кислородных вакансий. При этом диффузионный процесс, вообще говоря, протекает, однако его интенсивность и вероятность рекомбинации незначительны и вакансии остаются незаполненными. Отметим, что возможно удержание определенной части ионов электродом, т.е. поверхность электрода может выступать в некотором роде резервуаром для ионов кислорода.

Далее рассмотрим результаты трехмерного моделирования с учетом реальной кристаллической структуры оксида кремния. Кристаллическая решетка представлена на **рисунке 1.8**. Исследуемая структура представляет собой матрицу альфа-кварца между парой инертных электродов. Объем моделирования составляет 10 нм в глубину между электродами и 40 на 40 нм в области, периодически продолжающейся в направлениях, параллельных электродам. Нижний электрод остается заземленным, а напряжение на верхнем электроде со временем меняется.

Общее число атомов в рассматриваемой системе составляет около 10^6 , число ячеек сетки для уравнения Пуассона равно $7 \cdot 10^5$. При моделировании ионы и электроны движутся в матрице под действием электрического поля, взаимодействуя друг с другом и блокируя имеющиеся вакантные состояния.

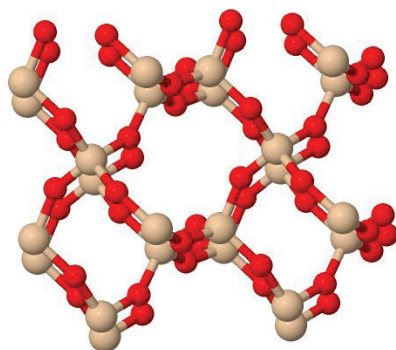


Рис. 1.8. Кристаллическая решетка SiO_2

Динамика ионов описывается в терминах переходов атомов кислорода между доступными состояниями в решетке. Кислородные узлы решетки альфа-кварца могут находиться в одном из трех состояний:

1. Кислородная вакансия с положительным зарядом;
2. Одиночный атом кислорода в собственном узле решетки;
3. Пара атомов кислорода с отрицательным зарядом, конкурирующая за узел решетки.

Система эволюционирует, атомы кислорода мигрируют между соседними узлами решетки, меняя свое состояние. Возможны четыре перехода ионов кислорода, меняющих состояния узлов:

1. Одиночный атом переходит в вакансию с соответствующим изменением состояния этих узлов;
2. Одиночный атом перескакивает к другому одиночному атому, образуя вакансию в первом узле и парное состояние во втором узле;

3. Один из атомов из парного состояния переходит в вакансию, образуя два одиночных состояния атома;

4. Один из атомов из парного состояния переходит к одиночному атому, меняя состояния соответствующих узлов.

Миграционные барьеры можно определить с помощью молекулярной динамики с потенциалом Терсоффа, который широко используется для моделирования систем на основе кремния [33, 34].

$$\left\{ \begin{array}{l}
 G = \sum G_i = \frac{1}{2} \sum_{i \neq j} V_{ij} \\
 V_{ij} = f_c(r_{ij}) (f_R(r_{ij}) + b_{ij} f_A(r_{ij})) \\
 f_R(r_{ij}) = A_{ij} \exp(-\lambda_{ij} r_{ij}) \\
 f_A(r_{ij}) = B_{ij} \exp(-\mu_{ij} r_{ij}) \\
 f_c(r_{ij}) = \begin{cases} 1, & r_{ij} < R_{ij} \\ \frac{1}{2} \left(1 + \cos \left(\pi \frac{r_{ij} - R_{ij}}{S_{ij} - R_{ij}} \right) \right), & R_{ij} \leq r_{ij} < S_{ij} \\ 0, & S_{ij} \leq r_{ij} \end{cases} \\
 b_{ij} = \chi_{ij} \left(1 + (\beta_i \zeta_{ij})^n \right)^{-1/2n_i} \\
 \zeta_{ij} = \sum_{k \neq i, j} f_c(r_{ik}) \omega_{ik} g(\theta_{ijk}) \\
 g(\theta_{ijk}) = 1 + \frac{c^2}{d^2} - \frac{c^2}{d^2 + (h_i - \cos \theta_{ijk})^2}
 \end{array} \right.$$

Здесь r_{ij} – расстояние между атомами i и j , а θ_{ijk} – валентный угол между связями ij и ik . Этот потенциал имеет 11 параметров для каждого типа атомов и по 2 параметра для каждой пары типов атомов, поэтому потенциал для системы с атомами двух сортов имеет всего 24 параметра. Параметры R , S и ω фиксируются, а остальные определяются в процедуре подгонки с использованием результатов квантово-механического моделирования.

При проведении расчетов мы использовали программное обеспечение с открытым исходным кодом Quantum Espresso для расчетов из первых принципов и собственный компьютерный код для моделирования молекулярной динамики. Псевдопотенциал и энергии отсечки были выбраны в соответствии со стандартной библиотекой твердотельных потенциалов с энергиями отсечки 60 Ry для волновых функций и 480 Ry для плотности

заряда. Суперячейка для вычисления энергии заряженных вакансий содержала 71 атом, и была принята сетка k-точек 5x5x4. В результате расчетная энергия образования ионов $O_2 - E_g$ составляет примерно 1 эВ. Энергетический барьер рекомбинации E_g в наших расчетах принимался равным нулю, так как переход из менее стабильного междоузлия в узел решетки гораздо более вероятен, чем из узла решетки в междоузлие.

Идентификация эмпирических параметров потенциала проводилась таким образом, чтобы они наилучшим образом соответствовали упругим свойствам, полученным из квантово-механического расчета. Чтобы начать процедуру подгонки, мы берем набор параметров из [34]. Результаты подгонки приведены в таблицах 1, 2. В таблице 1 показаны свойства упругости, полученные с помощью моделирования квантовой механики (КМ) и молекулярной динамики (МД). В табл. 2 представлены полученные параметры потенциала. Здесь мы также приводим параметры, взятые из [34].

Таблица 1

	<i>QM</i>	<i>MD</i>
B [GPa]	32	33.9
c11 [GPa]	82	94.3
c22 [GPa]	82	94.3
c33 [GPa]	90	61.1
c44 [GPa]	52	45.1
c55 [GPa]	52	40.4
c66 [GPa]	41	45.8
c12 [GPa]	-0.5	-0.49
c13 [GPa]	9	10.1
c23 [GPa]	9	7.6

Таблица 2

	Si [36]	O [36]	Si (present work)	O (present work)
<i>A</i>	1830.8	1885.5	1890.8	1922.8
<i>B</i>	471.18	218.79	460.97	214.20
<i>λ</i>	2.4799	4.1711	2.5714	4.2942
<i>μ</i>	1.7322	2.3569	1.7401	2.3726
<i>β</i>	1.1e-6	1.1632e-7	1.2716e-06	9.9807e-08
<i>n</i>	0.78734	1.04968	0.81812	1.10863
<i>c</i>	100390	64692.1	121907.3	53174.65
<i>d</i>	16.217	4.11127	16.005	4.375
<i>h</i>	-0.59825	-0.845922	-0.68726	-0.83058
<i>R</i>	2.5	1.7	2.5	1.7
<i>S</i>	2.8	2.0	2.8	2.0
<i>ω</i>		1		1
<i>χ</i>		1.17945		1.17506

Применяя метод эластичной ленты к переходу иона между уже занятыми состояниями, мы можем определить барьер для этого перехода. Для перехода из одного места в другое междоузельный ион O^{2-} должен разорвать связи с одним из атомов Si и вращаться вокруг другого, создавая вакансию, которая впоследствии может быть занята другим ионом O^{2-} или электроном. При вращении вокруг одного из атомов Si ион должен вытолкнуть другой атом кислорода из положения равновесия и создать связь с соответствующим атомом Si. Полная энергия системы будет изменяться во время миграции ионов за счет удлинения связей, стремящихся притянуть ион и свободный атом к их соответствующим положениям равновесия, но, когда ион переместится достаточно далеко, новое положение равновесия будет более предпочтительным, таким образом, система стабилизируется в новом

состоянии. Энергии системы в начале и в конце различны из-за разных углов между атомами.

Изменение энергии системы и процесс перехода показаны на **рисунках 1.9** и **1.10** соответственно.

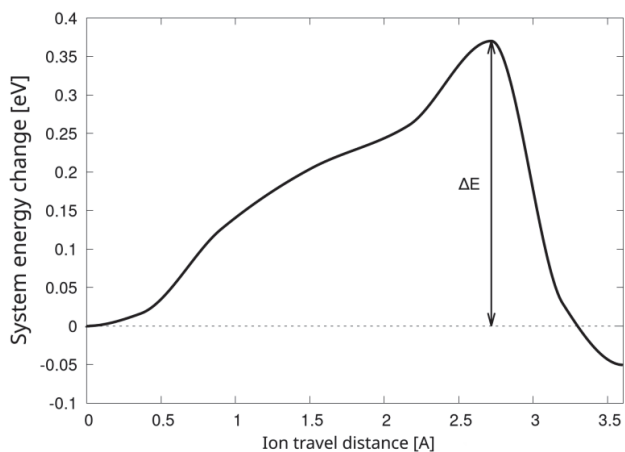


Рис. 1.9. Изменение энергии в методе эластичной ленты.

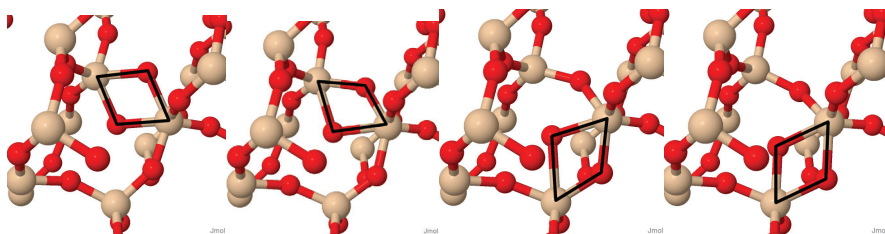


Рис. 1.10. Состояния решетки в эластичной ленте

Подведем итог вышеописанным шагам. Имея кристаллическую структуру материала, мы проводим квантово-механическое моделирование и извлекаем соответствующие данные, такие как энергия системы, упругие свойства материала и энергия образования дефектов. Энергия системы и константы упругости используются для идентификации межатомного потенциала Терсоффа. Этот потенциал используется в моделировании молекулярной динамики. С помощью молекулярно-динамического

моделирования определяются энергетические барьеры для миграции ионов. Полученные энергетические барьеры и энергия образования дефектов, извлеченная из результатов квантово-механического моделирования используются на следующем этапе Монте-Карло моделирования динамики всей системы.

Электроны считаются локализованными в ловушках. В альфа-кварце ловушки могут иметь двойное происхождение: вакантные кислородные узлы решетки и собственное расположение ловушек на некоторых атомах кремния. И, как и ионы, электроны могут перемещаться между этими ловушками, что осуществляется посредством квантового туннелирования.

Далее моделируется образование и разрушение филаментной структуры кинетическим методом Монте-Карло. Изначально матрица имеет небольшое количество дефектов, предположительно сосредоточенных вокруг границ зерен. Температура вокруг этой области фиксирована, так как трудно установить разумные граничные условия для температуры на электродах. На **рисунках 1.11, 1.12** представлена эволюция филамента. Вакансии кислорода показаны красным цветом, а ионы – синим цветом. Во время установки положительное напряжение подается на верхний электрод, а нижний электрод остается заземленным. Напряжение изменяется с постоянной скоростью 0,1 В/мкс в течение 60 мкс, пока не достигнет 6 В, затем приложенное напряжение снижается с той же скоростью, пока не достигнет нуля. На **рис. 1.11** показано формирование проводящего филамента при увеличении напряжения до 6В. Здесь представлено развитие процесса во времени. Текущие значения напряжения показаны в левом верхнем углу каждого рисунка. Последние два рисунка показывают, что филамент остается стабильным до подачи обратного напряжения. На **рис. 1.12** показано разрушение проводящего филамента под действием обратного напряжения. Филамент повреждается, что препятствует его проводимости, но не разрушается полностью, что облегчает последующие переключения.

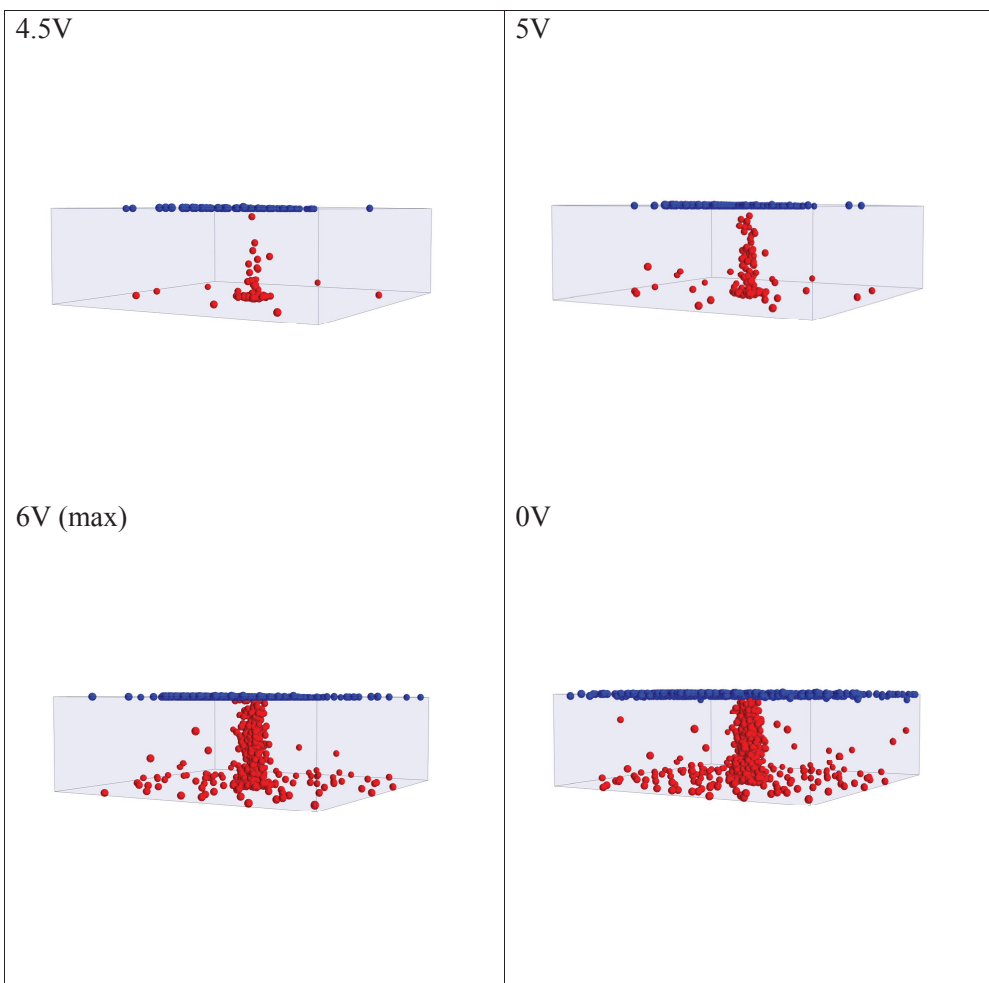


Рис. 1.11. Процесс образования филамента

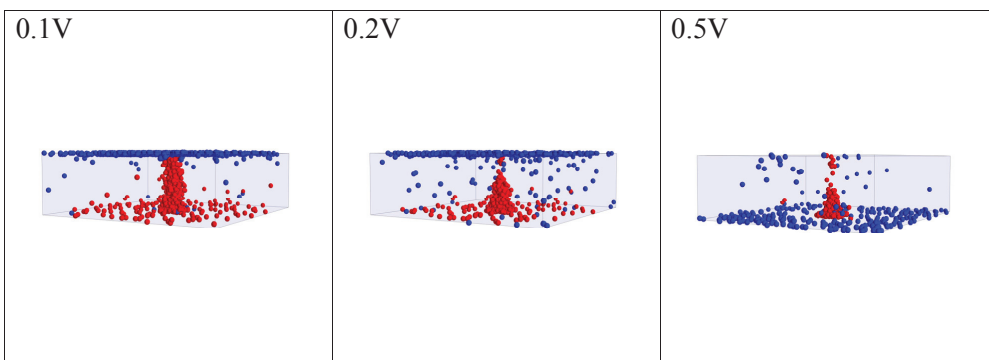


Рис. 1.12. Процесс разрушения филамента

1.5 Выводы

Реализована многомасштабная вычислительная схема динамики ионов в оксидных слоях. Многомасштабный подход позволяет изучать образование/разрушение филаментных структур в оксидных слоях на основе первичной информации – данных о химическом составе материала и его кристаллической структуре. На первом уровне проводится квантово-механическое моделирование, в результате чего определяются энергия активации кислородных вакансий, а также значения свойств материала, которые используются для построения межатомного потенциала. На втором уровне с помощью построенного потенциала осуществляется молекулярно-динамическое моделирование, целью которого является определение энергетических барьеров миграции ионов кислорода. На третьем уровне проводится моделирование динамики ионов на реальной кристаллической решетке методом Монте-Карло с использованием полученных миграционных барьеров. Результаты вычислительных экспериментов демонстрируют этапы роста и разрушения филаментных структур под действием переменного во времени напряжения.

Список источников к главе 1

1. Wong H.-S. P. et al., Metal-oxide RRAM. *Proceedings of the IEEE*, 2012. Vol. 100, no. 6, pp. 1951–1970.
2. Yang J. J., Strukov D. B., Stewart D. R., Memristive devices for computing. *Nature Nanotechnology*, 2013. Vol. 8, no. 1, pp. 13–24.
3. Li C., Hu M., Li Y., et al., Analogue signal and image processing with large memristor crossbars. *Nature Electronics*. Vol. 1, no. 1, Dec. 2017, pp. 52–59.
4. Mladenov V., Analysis and Simulations of Hybrid Memory Scheme Based on Memristors. *Electronics*, 2018, 7(11), 289.
5. Yang J. J., Pickett M. D., Xuema L., Ohlberg D. A. A., Stewart D. R., Williams R. S. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature nanotechnology*. Vol. 3, no. 7, 2008, p. 429.
6. Pickett M. D., Stukov D. B., Borghetti J. L., Yang J. J., Snider G. S., Stewart D. R., Williams R. S. Switching dynamics in titanium dioxide memristive devices. *Journal of Applied Physics*. Vol. 106, no. 7, 2009, p. 074508.
7. T. Heisig, K. Lange, A. Gutsche, K. Thorben Goß, S. Hamsch, A. Locatelli, T. Onur Menteş, F. Genuzio, S. Menzel, and R. Dittmann. Chemical Structure of Conductive Filaments in Tantalum Oxide Memristive Devices and Its Implications for the Formation Mechanism. *Adv. Electron.Mater.* 2022, 2100936
8. C. Funck and S. Menzel. Comprehensive Model of Electron Conduction in Oxide-Based Memristive Devices. *ACS Appl. Electron. Mater.* 2021, 3, 3674-3692
9. Tadmor, E.B.; Miller, R.E. *Modeling materials. Continuum, Atomistic and Multiscale Techniques*. Cambridge university press. 2014.
10. Steinhauser M. O. *Computational Multiscale Modeling of Fluids and Solids: Theory and Applications*, Springer, 2008.

11. W. Li, Y. Ando, E. Minamitani, S. Watanabe, Study of Li atom diffusion in amorphous Li_3PO_4 with neural network potential. *J. Chem. Phys.* 2017, 147, 214106.
12. Абгарян К.К. *Многомасштабное моделирование в задачах структурного материаловедения*. М: МАКС Пресс, 2017.
13. Zhuravlev A.A., Abgaryan K.K. Reviznikov, D.L. Multiscale discrete element modeling. *Symmetry*. 2021. 13 (2): 219. P. 1-11.
14. D. Ielmini and V. Milo, "Physics-based modeling approaches of resistive switching devices for memory and in-memory computing applications. *Journal of Computational Electronics*, vol. 16, no 4, pp. 1121–1143, November 2017.
15. Roldán, J.B.; González-Cordero, G.; Picos, R.; Miranda, E.; Palumbo, F.; Jiménez-Molinos, F.; Moreno, E.; Maldonado, D.; Baldomá, S.B.; Moner Al Chawa, M.; et al. On the Thermal Models for Resistive Random Access Memory Circuit Simulation. *Nanomaterials*, 2021, 11, 1261.
16. S. Kim, S.-J. Kim, K. M. Kim, S. R. Lee, M. Chang, E. Cho, Y.-B. Kim, C. J. Kim, U. -I. Chung, and I.-K. Yoo, "Physical electro-thermal model of resistive switching in bi-layered resistance-change memory. *Scientific Reports*, vol. 3, Article number 1680, pp. 1-6, April 2013.
17. Agudov N. V., Safonov A. V., Krichigin A. V., Kharcheva A. A., Dubkov A. A., Valenti D., Guseinov D. V., Belov A. I., Mikhaylov A. N., Carollo A., Spagnolo B., Nonstationary distributions and relaxation times in a stochastic model of memristor. *Journal of Statistical Mechanics: Theory and Experiment*, 2020, 024003.
18. T. Sadi, A. Mehonic, L. Montesi, M. Buckwell, A. Kenyon, A. Asenov, J. *Phys.: Condens. Matter*. 2018, 30, 084005.
19. D.V. Guseinov, D.S. Korolev, A.I. Belov, E.V. Okulich, V.I. Okulich, D.I. Tetelbaum, A.N. Mikhaylov. Flexible Monte-Carlo Approach to Simulate

- Electroforming and Resistive Switching in Filamentary Metal-Oxide Memristive Devices. *Modeling Simul. Mater. Sci. Eng.*, 2019. 28, 1, 017007.
20. S. Yu, X. Guan, and H.-S. P. Wong, "On the stochastic nature of resistive switching in metal oxide RRAM: Physical modeling, Monte Carlo simulation, and experimental characterization. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, December 2011, pp. 17.3.1-4.
 21. Biolek Z., Biolek D., Biolkova V. SPICE Model of Memristor with Nonlinear Dopant Drift. *Radioengineering*. Vol. 18, no. 2, 2009.
 12. Prodromakis T., Peh B. P., Papavassiliou C., Toumazou C. A versatile memristor model with nonlinear dopant kinetics. *IEEE Transactions on Electron Devices*. Vol. 58, no. 9, 2011, pp. 3099–3105.
 23. Kvatinsky S., Ramadan M., Friedman E. G., Kolodny A. VTEAM: A general model for voltage-controlled memristors. *IEEE Transactions on Circuits and Systems II: Express Briefs*. Vol. 62, no. 8, 2015, pp. 786–790.
 24. Zha J., Huang H., Liu Y. A novel window function for memristor model with application in programming analog circuits. *IEEE Transactions on Circuits and Systems II: Express Briefs*. Vol. 63, no. 5, 2015, pp. 423–427.
 25. Yakopcic C., Taha T. M., Subramanyam G., Pino R. E., Rogers S. A memristor device model. *IEEE electron device letters*. Vol. 32, no. 10, 2011, pp. 1436–1438.
 26. Yuehua Dai, Xiaoqing Wang, Bin Yang, An Improved Memristor Model Based on the Electrochemical Metallization Effect as a Synapse for Biomimetic Applications. *Physica Status Solidi (B)*. 2021. 259, 3.
 27. Qunhai Sun, Yuehua Dai, Modeling and Synaptic behaviors of a WO_x-based Memristor with Oxygen Vacancy Conductive Mechanism. *Physica Status Solidi (B)*. 2022. 2100647.

28. Morozov A. Yu., Abgaryan K. K., Reviznikov D. L. Mathematical model of a neuromorphic network based on memristive elements. *Chaos, Solitons & Fractals*. 2021. V. 143. P. 110548.
29. Morozov A. Yu., Abgaryan K. K., Reviznikov D. L. Mathematical Modeling of an Analogue Self-Learning Neural Network Based on Memristive Elements Taking into Account Stochastic Switching Dynamics. *Nanobiotechnology Reports*, 2021, Vol. 16, No. 6, pp. 767–776.
30. D. Gao, A. El-Sayed, A. Shluger, A mechanism for Frenkel defect creation in amorphous SiO₂ facilitated by electron injection. *Nanotechnology*. 2016, 27, 505207.
31. D. Gao, J. Strand, M. Munde, A. Shluger, Mechanisms of Oxygen Vacancy Aggregation in SiO₂ and HfO₂. *Front. Phys.* 2019, 7, 43.
32. G. Mills, H. Jonsson, Quantum and thermal effects in H₂ dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Phys. Rev. Lett.* 1994, 72, 1124.
33. Tersoff, J. New empirical approach for the structure and energy of covalent systems. *Physical Review B*. 1988, 37(12), 6991-7000.
34. S. Munetoh, T. Motooka, K. Moriguchi, A. Shintani, Interatomic potential for Si-O systems using Tersoff parameterization. *Comp. Mater. Sci.* 2007, 39, 334.

ГЛАВА 2

ДИНАМИЧЕСКИЕ МОДЕЛИ МЕМРИСТОРА

Мемристор является относительно новым электрическим элементом и представляет собой резистор, проводимость которого меняется в зависимости от суммарного протекшего через него электрического заряда. При этом установившееся сопротивление в отсутствие тока не изменяется со временем. То есть мемристор – это элементарная ячейка долгосрочной энергонезависимой памяти [1, 2].

Мемристивный эффект, как правило, возникает за счет перемещения ионов в сверхтонком диэлектрическом слое при приложении электрического поля. Применительно к различным оксидам часто говорят о перемещении вакансий кислорода и формировании/разрушении проводящих филаментов. Большинство известных моделей мемристора сформулированы в виде динамической системы относительно состояния мемристора. Параметр состояния мемристора – это величина, которая соответствует положению границы, разделяющей области с низкой и высокой концентрацией вакансий кислорода, толщине проводящего слоя, или толщине непроводящего барьера, в котором возникает туннельный ток электронов. В зависимости от закона изменения параметра состояния мемристора можно выделить несколько математических моделей – в частности, модели линейного [3] и нелинейного дрейфа [4], модель, основанная на барьере Симмонса [5]. Для ограничения переменной состояния вводятся специальные функции окна [6–9]. Экспериментальные данные показывают, что изменение состояния происходит не при любом значении напряжения, а начиная с некоторого порога, в связи с этим в модели добавляются пороговые условия [10–12].

Согласно экспериментальным данным [13, 14], характер функционирования мемристоров является отчасти стохастическим, о чем

свидетельствует разброс в вольт-амперных характеристиках в процессе переключения из высокоомного состояния в низкоомное состояние и наоборот. Для учета этой особенности рассматривается два подхода к модификации существующих моделей мемристоров. Первый подход заключается в замене детерминированного уравнения состояния на стохастическое путем введения слагаемого, отвечающего за аддитивный (гауссовский) шум. А второй подход подразумевает использование интервального аппарата: идея заключается во введении в математическую модель интервальных параметров. В этом случае результатом имитационного моделирования в каждый момент времени будут являться интервальные оценки фазовых переменных. Значения интервальных параметров подбираются таким образом, что полученные оценки полностью содержат экспериментальные данные.

В разделе 2.1 рассматриваются две детерминированные модели мемристора: первая, основанная на экспоненциальной модели дрейфа легирующей примеси и вторая с нелинейной зависимостью от напряжения. Последующие разделы посвящены вопросам учета стохастических особенностей функционирования мемристоров. В разделе 2.2 выполняется модификация моделей путем замены детерминированного уравнения состояния на стохастическое уравнение, а в разделе 2.3 в модели вводятся интервальные параметры.

2.1. Представление мемристора динамической системой.

Детерминированные модели

Рассматриваются несколько моделей мемристоров. Первая модель является переменнорезисторной моделью тонкопленочного мемристора, основанной на экспоненциальной модели дрейфа легирующей примеси [15]:

$$\frac{dx(t)}{dt} = \begin{cases} \mu_v \frac{V_p}{D^2} \exp\left(\frac{R_{on}}{V_p} I(t)\right), & V(t) \geq V_p, \\ \mu_v \frac{V_n}{D^2} \exp\left(\frac{R_{on}}{V_n} I(t)\right), & V(t) \leq V_n, \\ \mu_v \frac{R_{on}}{D^2} I(t), & V_n < V(t) < V_p, \end{cases}$$

$$R(t) = R_{on} x(t) + R_{off} (1 - x(t)), \quad (2.1)$$

$$I(t) = \frac{V(t)}{R(t)},$$

где $x(t) \in [0, 1]$ – переменная состояния; R_{on} , R_{off} – минимальное и максимальное сопротивление мемристора; $I(t)$, $V(t)$, $R(t)$ – значение тока, напряжения и сопротивления мемристора в момент времени t ; V_p , V_n – значения напряжений, при которых происходит переключение состояния; μ_v – коэффициент легирующей подвижности; D – толщина полупроводниковой пленки.

Выполняется моделирование работы мемристора при следующих значениях параметров: $R_{on} = 205$ Ом, $R_{off} = 2.13$ кОм, $\mu_v = 6 \times 10^{-10}$, $V_p = 0.65$ В, $V_n = -0.87$ В, $D = 620$ нм, $x(0) = 0.1$, $t \in [0, 16]$ мс, $V(t)$ – рис. 2.16. Такой выбор параметров и формы напряжения $V(t)$ обусловлен получением характеристик мемристора, схожих с экспериментальными характеристиками по оксиду титана, приведенными в работе [3].

На рис. 2.1а показана экспериментальная вольтамперная характеристика и модельная.

Здесь наблюдается хорошее согласование в правой части графиков и удовлетворительное – в левой части.

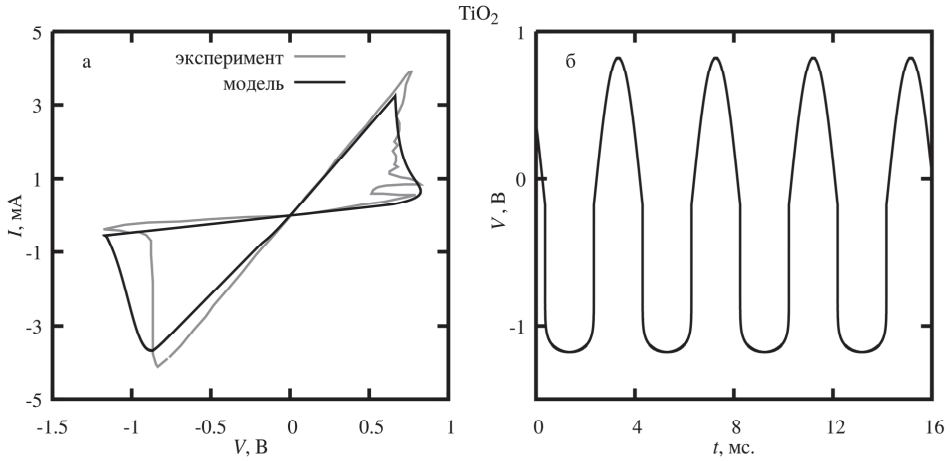


Рис. 2.1. Сравнение вольт-амперной характеристики модели (2.1) с экспериментальными данными по оксиду титана (а) при определенной форме входного напряжения (б)

Далее рассматривается модель с нелинейной зависимостью от напряжения. В общем виде уравнение, описывающее состояние мемристора, может быть представлено следующим образом:

$$\frac{dx(t)}{dt} = a \cdot f(x(t)) \cdot V(t)^s,$$

где $x(t) \in [0, 1]$ – переменная состояния, a – некоторая константа, s – нечетное целое число, f – функция окна, используемая для приблизительного представления нелинейных эффектов ионного дрейфа и ограничения границ. Часто используется оконная функция Вioлек [7]:

$$f(x, V) = \begin{cases} 1 - (x - 1)^{2p}, & V \leq 0, \\ 1 - x^{2p}, & V > 0. \end{cases} \quad (2.2)$$

В работе используется модель мемристора, предложенная в [16]:

$$\frac{dx(t)}{dt} = a \cdot V(t)^s \begin{cases} 1 - (1 - x(t))^{2 \operatorname{round}\left(\frac{b}{|V(t)|+c}\right)}, & V(t) \leq -V_{thr}, \\ 1 - x(t)^{2 \operatorname{round}\left(\frac{b}{|V(t)|+c}\right)}, & V(t) > V_{thr}, \\ 0, & -V_{thr} < V(t) \leq V_{thr}, \end{cases}$$

$$I(t) = x(t)^n \beta \sinh(\alpha_M V(t)) + \chi [\exp(\gamma V(t)) - 1], \quad (2.3)$$

$$R(t) = \frac{V(t)}{I(t)},$$

где $x(t) \in [0, 1]$ – переменная состояния; $I(t)$, $V(t)$, $R(t)$ – значение тока, напряжения и сопротивления мемристора в момент времени t ; V_{thr} – пороговое значение напряжения активации; a – константа; s – нечетное целое число; $n, \beta, \alpha_M, \chi, \gamma$ – подгоночные параметры функции тока; round – функция получения целочисленного результата; b, c – подгоночные коэффициенты. Несложно убедиться в схожести оконной функции, используемой в модели с оконной функцией (2.2).

Выполняется моделирование работы мемристора при следующих значениях параметров: $n = 5$, $\beta = 7.069 \cdot 10^{-5}$, $\alpha_M = 1.8$, $\chi = 1.946 \cdot 10^{-4}$, $\gamma = 0.15$, $a = 1$, $s = 5$, $b = 15$, $c = 2$, $V_{thr} = 1$, $x(0) = 0.4$. Часть значений ($\alpha_M, \gamma, a, s, b, c, V_{thr}, x(0)$) соответствует значениям в исходной работе [16], а часть подобрана для наибольшего соответствия экспериментальным данным по оксиду гафния (HfO_2) приведенным в той же статье [16].

На рис. 2.2а приведено сравнение полученной вольт-амперной характеристики с экспериментальной кривой для HfO_2 . Зависимость напряжения от времени (рис. 2.2б) взята из [16].

Здесь наблюдается удовлетворительное согласование с экспериментальными данными.

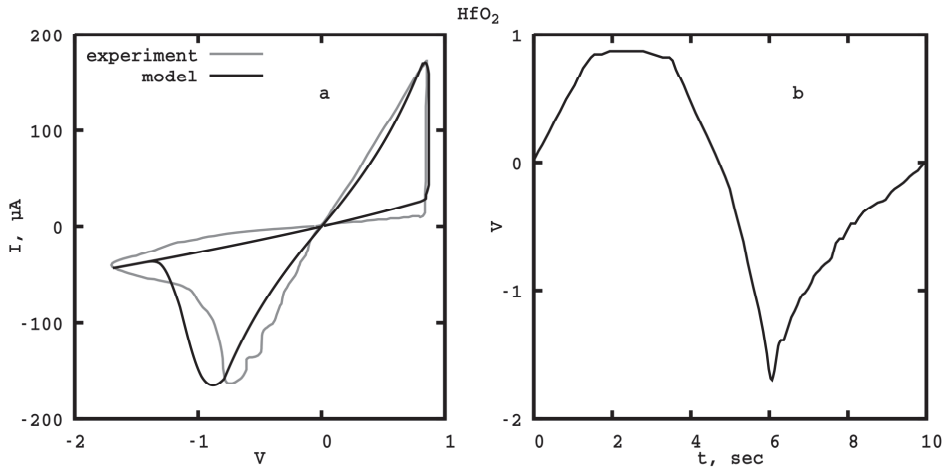


Рис. 2.2. Сравнение вольт-амперной характеристики модели (2.3) с экспериментальными данными по оксиду гафния (а) при определенной форме входного напряжения (б)

На рис. 2.3а представлена модельная ВАХ полученная в случае синусоидальной зависимости напряжения от времени (рис. 2.3б).

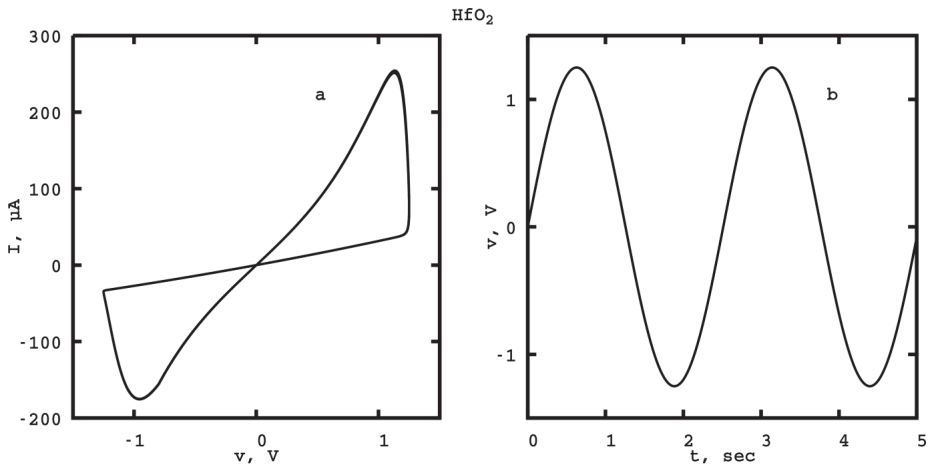


Рис. 2.3. Вольт-амперная характеристика модели (2.3) (а) при синусоидальной форме входного напряжения (б)

Данная модель реализует мемристорную модель Лехтонена-Лайхо, которая основана на физических экспериментах и на механизме протекания тока через полупроводниковые аморфные оксиды переходных металлов,

обладает хорошей точностью и дает возможность тонкой настройки. Оконная функция, используемая в (2.3), позволяет применять модель для различных напряжений, что продемонстрировано результатами расчетов.

Без ограничения общности можно предположить, что модель мемристора представляется в виде двух уравнений: первое определяет зависимость скорости изменения состояния мемристора от тока ($I(t)$), приложенного напряжения ($V(t)$) и состояния ($x(t)$), а второе определяет величину сопротивления мемристора:

$$\begin{aligned}\frac{dx(t)}{dt} &= F_x(I(t), V(t), x(t)), \\ R(t) &= F_R(V(t), x(t))\end{aligned}$$

В главе 3 при описании модели нейроморфной сети будут использоваться введенные обозначения F_x и F_R .

В следующих разделах рассматриваются различные модификации описанных моделей для учета стохастических особенностей функционирования мемристивных элементов.

2.2. Стохастические модели мемристоров

Одним из важных вопросов свойств мемристоров является стохастическая динамика переключения [14], которая в основном возникает в результате хаотического движения ионов (или кислородных вакансий) в диэлектрической пленке под действием электрического поля. Подробное экспериментальное и теоретическое исследование этого эффекта проведено в работах [17], [18] и [19]. Было отмечено, что внутренние и внешние шумы могут играть положительную роль в процессе резистивного переключения, вызывая рост контрастности между состояниями с низким и высоким сопротивлением, а также ускорение процесса релаксации. Эти эффекты можно рассматривать как явления нелинейной релаксации в метастабильных системах конденсированного состояния при шумовом воздействии [20], [21].

Основные подходы к учету стохастических эффектов в математической модели описаны в статье [17]. Одним из самых простых способов является использование стохастического дифференциального уравнения для переменной состояния вместо детерминированного. В [22] авторы использовали уравнение состояния с шумом $dx = -\nabla V(x)dt + G(t)dW$ (где x – вектор переменных стохастического процесса; $V(x)$ – присущий системе потенциал; W – винеровский процесс; $G(t)$ – функция, задающая сигнал, поступающий в систему) и наблюдали эффект Стохастический резонанс в мемристоре. Более полная модель, основанная на одномерном уравнении дрейфа-диффузии для концентрации кислородных вакансий в мемристорном материале, предложена в [17]. Модель позволила авторам тщательно исследовать динамику переключения мемристоров с учетом внутренних и внешних шумов.

В соответствии с работой [22] учет стохастических особенностей будет выполняться за счет введения в обыкновенное дифференциальное уравнение состояния стохастической добавки в виде аддитивного белого (гауссова) шума. Уравнение, которое задает изменение переменной состояния мемристора, можно записать следующим образом:

$$dx(t) = F(x(t))dt + \eta dW,$$

где F – определяет скорость изменения переменной состояния; η – коэффициент характеризующий интенсивность шума; W – винеровский процесс.

Для получения приближенной реализации стохастического процесса $x(t)$ используется метод Рунге-Кутты порядка 1.5. Определим приближенное решение на сетке $t_0 < t_1 < t_2 < \dots < t_{N-1} < t_N$ при начальном условии $x(t_0) = x_0$:

$$\begin{aligned} x_k^1 &= x_k + \frac{2}{3} [F(x_k) \Delta t_{k+1} + \eta \Delta W_{k+1}], \\ x_k^2 &= x_k + [F(x_k^1) - F(x_k)] \Delta t_{k+1}, \\ x_{k+1} &= x_k + \frac{1}{4} [3F(x_k^1) + F(x_k^2)] \Delta t_{k+1} + \eta \Delta W_{k+1}, \end{aligned}$$

где $\Delta t_{k+1} = t_{k+1} - t_k$, $\Delta W_{k+1} = W(t_{k+1}) - W(t_k)$. Пусть $N(0,1)$ нормально распределенная случайная величина с нулевым математическим ожиданием и единичной дисперсией. Тогда случайное значение ΔW_k вычисляется как

$$\Delta W_k = z_k \sqrt{\Delta t_k},$$

где z_k выбирается из $N(0,1)$.

Отметим, что можно использовать другие методы интегрирования стохастических ОДУ, например метод Мильштейна первого порядка, который для данной задачи будет эквивалентен методу Эйлера-Маруямы. В этом случае решение на $(k+1)$ временном слое будет определяться как

$$x_{k+1} = x_k + F(x_k) \Delta t_{k+1} + \eta \Delta W_{k+1}.$$

Модифицированная модель мемристора (2.1) с аддитивным шумом описывается следующими уравнениями:

$$dx(t) = \eta dW + \begin{cases} \mu_v \frac{V_p}{D^2} \exp\left(\frac{R_{on}}{V_p} I(t)\right) dt, & V(t) \geq V_p, \\ \mu_v \frac{V_n}{D^2} \exp\left(\frac{R_{on}}{V_n} I(t)\right) dt, & V(t) \leq V_n, \\ \mu_v \frac{R_{on}}{D^2} I(t) dt, & V_n < V(t) < V_p, \end{cases}$$

$$R(t) = R_{on} x(t) + R_{off} (1 - x(t)), \quad (2.4)$$

$$I(t) = \frac{V(t)}{R(t)},$$

где $x(t) \in [0,1]$ – переменная состояния; R_{on} , R_{off} – минимальное и максимальное сопротивление мемристора; $I(t)$, $V(t)$, $R(t)$ – значение тока, напряжения и сопротивления мемристора в момент времени t ; V_p , V_n – значения напряжений, при которых происходит переключение состояния; μ_v – коэффициент легирующей подвижности; D – толщина полупроводниковой пленки; η – коэффициент, характеризующий интенсивность шума; W – винеровский процесс.

Выполняется моделирование работы мемристора при следующих значениях параметров: $\eta = 0.05$, $R_{on} = 25$ кОм, $R_{off} = 250$ кОм, $\mu_v = 10^{-14}$, $V_p = 1.5$ В, $V_n = -1$ В, $D = 20$ нм, $x(0) = 0$, $V(t)$ – рис. 2.4а. Такой выбор параметров и формы напряжения обусловлен получением характеристик мемристора, схожих с экспериментальными характеристиками по оксиду титана, приведенными в работе [3]. Зависимость тока от времени показана на рис. 2.4б, сопротивления на рис. 2.4с и изменение состояния на рис. 2.4д. На рис. 2.5 показана экспериментальная вольтамперная характеристика и модельная.

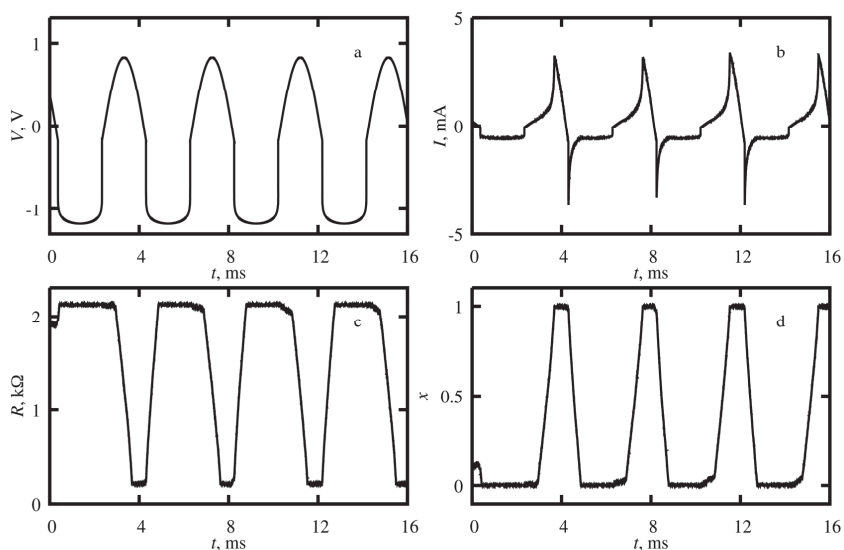


Рис. 2.4. Зависимость напряжения (а), тока (б), сопротивления (с) и состояния (д) от времени в процессе моделирования мемристора (2.4)

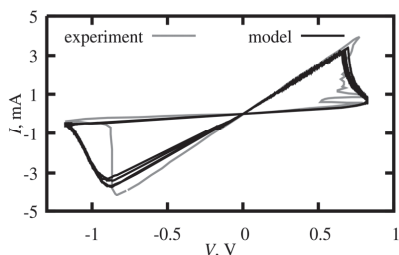


Рис. 2.5. Сравнение вольт-амперной характеристики, полученной в процессе моделирования мемристора (2.4), с экспериментальными данными по оксиду титана

Наличие шума в модели мемристора приводит к тому, что все параметры приобретают стохастические свойства. На графике ВАХа различимы траектории, которые соответствуют разным циклам переключения мемристора. Здесь наблюдается хорошее согласование в правой части графиков и удовлетворительное – в левой части.

Далее рассмотрим модифицированную модель мемристора (2.3) с аддитивным шумом:

$$dx(t) = \eta dW + a \cdot V(t)^s \begin{cases} \left[1 - (1-x)^{2 \text{round}\left(\frac{b}{|V(t)|+c}\right)} \right] dt, V(t) \leq -V_{thr}, \\ \left[1 - x^{2 \text{round}\left(\frac{b}{|V(t)|+c}\right)} \right] dt, V(t) > V_{thr}, \\ 0, -V_{thr} < V(t) \leq V_{thr}, \end{cases}$$

$$I(t) = x(t)^n \beta \sinh(\alpha_M V(t)) + \chi [\exp(\gamma V(t)) - 1], \quad (2.5)$$

$$R(t) = \frac{V(t)}{I(t)},$$

где $x(t) \in [0, 1]$ – переменная состояния; $I(t)$, $V(t)$, $R(t)$ – значение тока, напряжения и сопротивления мемристора в момент времени t ; V_{thr} – пороговое значение напряжения активации; a – константа; s – нечетное целое число; $n, \beta, \alpha_M, \chi, \gamma$ – подгоночные параметры функции тока; round – функция получения целочисленного результата; b, c – подгоночные коэффициенты; η – коэффициент, характеризующий интенсивность шума; W – винеровский процесс.

Выполняется моделирование работы мемристора при следующих значениях параметров: $n=5$, $\beta=7.069 \cdot 10^{-5}$ А, $\alpha_M=1.8$ В⁻¹, $\chi=1.946 \cdot 10^{-4}$ А, $\gamma=0.15$ В⁻¹, $a=1$, В⁻⁵, $s=5$, $b=15$ В, $c=2$ В, $v_{thr}=1$ В, $x(0)=0$, $V(t)$ – рис. 2.6, а. Часть значений ($\alpha_M, \gamma, a, s, b, c, v_{thr}$) соответствует значениям, приведенным в исходной работе [16], а часть (β, χ) подобрана для

наилучшего соответствия экспериментальным данным [16] по оксиду гафния (HfO_2) в рамках детерминированной модели ($\eta = 0$).

На рис. 2.6, b, c и d показано сравнение модельной вольтамперной характеристики при различных значениях коэффициента шума с экспериментальной кривой из работы [16].

На графиках ВАХ присутствует несколько случайных траекторий, каждая из которых соответствует определенному циклу переключения мемристора. Следует отметить, что полученный разброс в вольт-амперных характеристиках на разных циклах согласуется с экспериментальными данными [13] для оксида гафния.

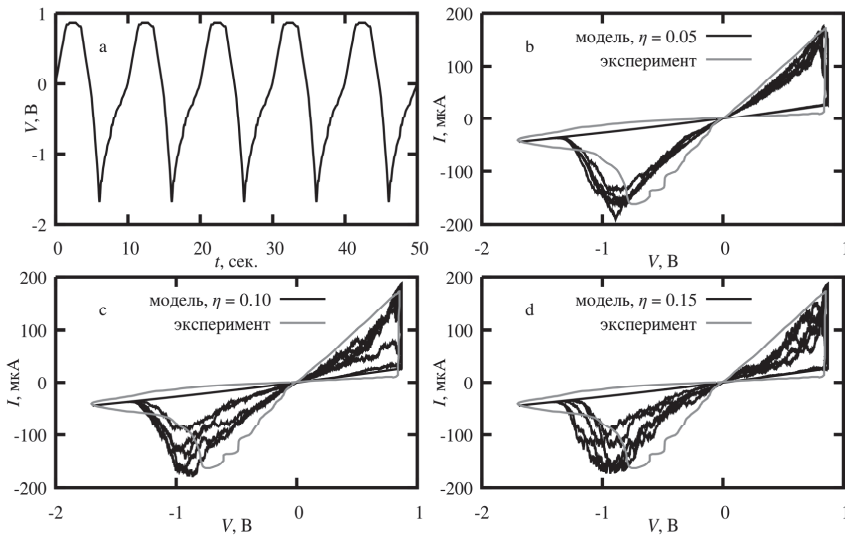


Рис. 2.6. Зависимость напряжения от времени (a), ВАХ при значении коэффициента шума $\eta = 0.05$ (b), $\eta = 0.10$ (c) и $\eta = 0.15$ (d) полученная в процессе моделирования мемристора (2.5)

Наличие шума в модели мемристора приводит к тому, что все параметры приобретают стохастические свойства. В частности при $\eta = 0.05$ оценка математического ожидания минимального сопротивления мемристора $M[R_{on}] \approx 6.77 \text{ кОм}$, а дисперсии $D[R_{on}] \approx 1 \text{ кОм}^2$. На другие параметры, такие как максимальное значение сопротивления и напряжения переключения

аддитивный шум сказывается в меньшей степени. Ясно, что с увеличением значения коэффициента шума, разброс параметров будет также увеличиваться.

2.3. Интервальные модели мемристоров

В данном разделе предлагается использовать интервальный аппарат для учета разброса характеристик мемристивных элементов.

Математические модели мемристоров традиционно формулируются в виде динамических систем относительно параметра состояния мемристора, которое характеризует уровень проводимости элемента. При этом динамические системы содержат набор параметров, значения которых подбираются из соображений наилучшего воспроизведения экспериментальных данных. Здесь, в частности, могут применяться методы, минимизирующие среднеквадратическое отклонение от эксперимента.

Исторически интервальные методы возникли в связи с потребностью в гарантированных вычислениях, которые учитывали бы погрешность самих вычислительных схем, а также ошибки округления при расчетах на ЭВМ. В настоящее же время интерес представляют задачи, в которых интервальность возникает непосредственно в самой постановке. В случае с интервальными динамическими системами результатом их моделирования будут являться интервальные оценки фазовых переменных в зависимости от времени.

Для моделирования динамических систем с интервальными параметрами ранее был разработан алгоритм адаптивной интерполяции [23], который относится к методам, представляющим решение в виде полинома относительно интервальных параметров [24, 25]. Его суть заключается в построении кусочно-полиномиальной функции, которая интерполирует зависимость решения задачи от точечных значений интервальных параметров. Алгоритм имеет теоретическое обоснование [23] и успешно применен к жестким системам обыкновенных дифференциальных уравнений (ОДУ), системам с динамическим хаосом и бифуркациями [26], прикладным

задачам химической кинетики и молекулярной динамики [27]. В работе [27] показана возможность использования разреженных сеток [28–30] в алгоритме адаптивной интерполяции для моделирования динамических систем, содержащих большое количество интервальных параметров, и продемонстрирована эффективность данного подхода.

Интервальная параметрическая идентификация заключается в нахождении таких интервальных значений параметров, при которых соответствующие интервальные оценки фазовых переменных будут полностью содержать экспериментальные данные (или минимизировать отклонение от них). В работе [31] разработан аппарат для интервальной параметрической идентификации динамических систем на основе алгоритма адаптивной интерполяции.

Сначала рассмотрим модификацию модели (2.1) с ограничением по максимальному току:

$$\frac{dx(t)}{dt} = \begin{cases} \mu_v \frac{V_p}{D^2} \exp\left(\frac{R_{on}}{V_p} I(t)\right), & V(t) \geq V_p, \\ \mu_v \frac{V_n}{D^2} \exp\left(\frac{R_{on}}{V_n} I(t)\right), & V(t) \leq V_n, \\ \mu_v \frac{R_{on}}{D^2} I(t), & V_n < V(t) < V_p, \end{cases}$$

$$R(t) = R_{on}x(t) + R_{off}(1 - x(t)),$$

$$I(t) = \min \left[I_{max}, \max \left[-I_{max}, \frac{V(t)}{R_{on}x(t) + R_{off}(1 - x(t))} \right] \right], \quad (2.6)$$

где $x \in [0, 1]$ – переменная состояния; $I(t)$, $V(t)$, $R(t)$ – значения тока, напряжения и сопротивления в момент времени t ; R_{on} , R_{off} – минимальное и максимальное сопротивление мемристора; V_n , V_p – значения напряжений, при которых происходит переключение состояния; μ_v – коэффициент легирующей подвижности; D – толщина полупроводниковой пленки; I_{max} – максимальное значение тока.

На рис. 2.7 показана экспериментальная зависимость тока от времени при определенной форме напряжения по мемристору на основе LiNbO_3 [32, 33]. Отметим, что типична ситуация, когда экспериментальные данные представляют собой пучок траекторий, который соответствует разным циклам переключения мемристора. Согласно [32], уровень напряжения в эксперименте изменялся с шагом 0,1 В каждые 50 мс. Всего было выполнено 30 циклов: $V: 0 \rightarrow +5 \rightarrow -5 \rightarrow 0$ В.

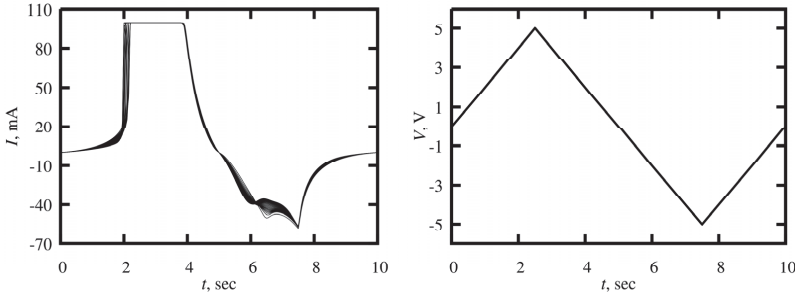


Рис. 2.7. Экспериментальная зависимость тока в 30 различных циклах переключения мемристора от времени (слева) и для заданной конкретной формы напряжения, приложенного к элементу (справа)

Выполним моделирование работы мемристора (2.6) при следующих значениях параметров: $R_{on} = 26.2$ Ом, $R_{off} = 211$ Ом, $V_n = -4.87$ В, $V_p = 2.9$ В, $\mu_v = 4.8 \times 10^{-14}$, $D = 4 \times 10^{-7}$ м, $I_{max} = 100$ мА, $x(0) = 0$, полученных с помощью классических подходов минимизирующих среднеквадратическое отклонение модели от эксперимента. На рис. 2.8 показан результат моделирования.

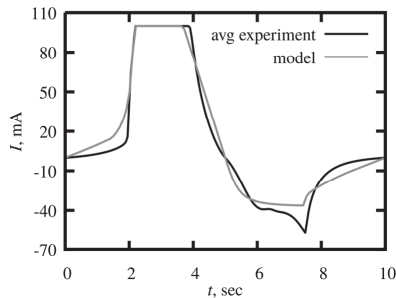


Рис. 2.8. Сравнение модельной кривой зависимости тока от времени (серый цвет) с усредненной экспериментальной кривой за 30 циклов (черный цвет):

$$I_{avg}(t) = \frac{1}{30} \sum_{i=1}^{30} I_i(t), \text{ где } I_i(t) \text{ – кривая зависимости тока от времени в } i\text{-ом цикле}$$

Здесь наблюдается удовлетворительное согласование модели с экспериментом. Модельная кривая аппроксимирует усредненную экспериментальную кривую, но не совпадает с ней.

Далее рассмотрим интервальную постановку задачи параметрической идентификации. Введем в модель несколько интервальных параметров $R_{on} \in [\underline{R}_{on}, \overline{R}_{on}]$, $R_{off} \in [\underline{R}_{off}, \overline{R}_{off}]$, $V_n \in [\underline{V}_n, \overline{V}_n]$, $V_p \in [\underline{V}_p, \overline{V}_p]$, $\mu_v \in [\underline{\mu}_v, \overline{\mu}_v]$, $\delta \in [\underline{\delta}, \overline{\delta}]$ где $\underline{R}_{on} \leq \overline{R}_{on}$, $\underline{R}_{off} \leq \overline{R}_{off}$, $\underline{V}_n \leq \overline{V}_n$, $\underline{V}_p \leq \overline{V}_p$, $\underline{\mu}_v \leq \overline{\mu}_v$, $\underline{\delta} \leq \overline{\delta}$ – нижние и верхние границы интервалов. Параметр δ введен в виде слагаемого с весовым коэффициентом, зависящим от напряжения, в уравнение тока для расширения обобщающих способностей модели. Обозначим через χ множество неопределенностей параметров:

$$\chi = [\underline{R}_{on}, \overline{R}_{on}] \times [\underline{R}_{off}, \overline{R}_{off}] \times [\underline{V}_n, \overline{V}_n] \times [\underline{V}_p, \overline{V}_p] \times [\underline{\mu}_v, \overline{\mu}_v] \times [\underline{\delta}, \overline{\delta}].$$

В данном случае состояние x и ток I в каждый момент времени t_k будут являться параметрическими множествами:

$$\begin{aligned} \mathbf{X}_k &= \left\{ x_k (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) \mid (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)^T \in \chi \right\}, \\ \mathbf{I}_k &= \left\{ I_k (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) \mid (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)^T \in \chi \right\}, \end{aligned}$$

где $x_k (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) = x(t_k)$ решение следующей системы ОДУ:

$$\begin{aligned} \frac{dx(t)}{dt} &= 6.25 \times 10^{12} \mu_v \begin{cases} V_p \exp\left(\frac{R_{on}}{V_p} I(t)\right), & V(t) \geq V_p, \\ R_{on} I(t), & V_n < V(t) < V_p, \\ V_n \exp\left(\frac{R_{on}}{V_n} I(t)\right), & V(t) \leq V_n, \end{cases} \\ I(t) &= \min \left[\frac{1}{10}, \max \left[-\frac{1}{10}, \frac{V(t)}{R_{on} x(t) + R_{off} (1-x(t))} + \left(1 + \frac{V(t)}{10}\right) \delta \right] \right] \\ x(0) &= 0, \quad t \in [0, t_k], \end{aligned}$$

$$I_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) = \min \left[\frac{1}{10}, \max \left[-\frac{1}{10}, \frac{V(t_k)}{R_{on}x_k + R_{off}(1-x_k)} + \left(1 + \frac{V(t_k)}{10} \right) \delta \right] \right].$$

Отметим, что важную роль играет явное представление $x_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$, которое можно получить с помощью алгоритма адаптивной интерполяции.

Обозначим, через \hat{I}_k усредненное экспериментальное значение тока в момент времени t_k . Задача интервальной параметрической идентификации заключается в нахождении таких $\underline{R}_{on}, \overline{R}_{on}, \underline{R}_{off}, \overline{R}_{off}, \underline{V}_n, \overline{V}_n, \underline{V}_p, \overline{V}_p, \underline{\mu}_v, \overline{\mu}_v, \underline{\delta}, \overline{\delta}$, при которых $\hat{I}_k \in \mathbf{I}_k$ для каждого k . Определение границ интервалов сводится к минимизации следующего функционала:

$$J(\underline{R}_{on}, \overline{R}_{on}, \underline{R}_{off}, \overline{R}_{off}, \underline{V}_n, \overline{V}_n, \underline{V}_p, \overline{V}_p, \underline{\mu}_v, \overline{\mu}_v, \underline{\delta}, \overline{\delta}) = \sum_{I_k \in \mathbf{I}_k} \min (I_k - \hat{I}_k)^2.$$

При значениях $\underline{R}_{on} = 14.46$, $\overline{R}_{on} = 29.46$, $\underline{R}_{off} = 209.6$, $\overline{R}_{off} = 210.8$, $\underline{V}_n = -5.2$, $\overline{V}_n = -4.66$, $\underline{V}_p = 2.88$, $\overline{V}_p = 3.04$, $\underline{\mu}_v = 3.77 \times 10^{-14}$, $\overline{\mu}_v = 5.03 \times 10^{-14}$, $\underline{\delta} = -0.01$, $\overline{\delta} = 0.012$, функционал будет равен нулю (при вычислении было взято 50 равномерно распределенных точек на интервале интегрирования, $k = 1, \dots, 50$). Это говорит о том, что каждое значение \hat{I}_k принадлежит соответствующему множеству \mathbf{I}_k , таким образом, модельное решение полностью покрывает усредненные экспериментальные данные.

На рис. 2.9 показана интервальная оценка зависимости тока от времени, а на рис. 2.10 – интервальная ВАХ, полученные при соответствующих интервальных значениях параметров $R_{on}, R_{off}, V_n, V_p, \mu_v, \delta$.

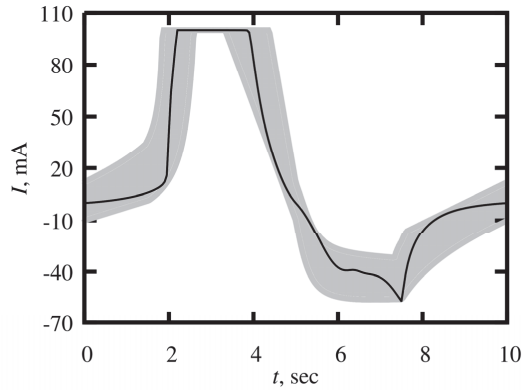


Рис. 2.9. Сравнение интервальной оценки зависимости тока от времени (серый цвет) с экспериментальной кривой (черный цвет)

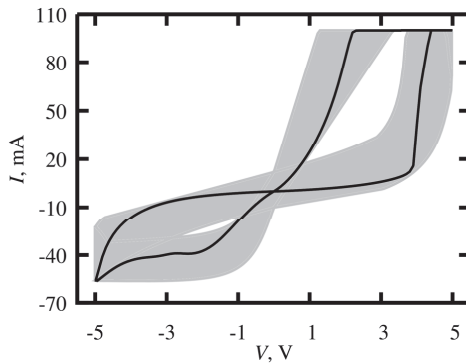


Рис. 2.10. Сравнение интервальной ВАХ (серый цвет) с экспериментальной кривой (черный цвет)

Интервальные оценки содержат в себе все экспериментальные данные. Однако необходимо отметить, что подобранные интервальные параметры могут быть неоптимальными с точки зрения ширины получаемых результатов. Может быть получен другой набор интервальных параметров, при котором интервальные оценки все так же будут содержать экспериментальные данные, но при этом они будут более узкими.

Далее приведем краткое описание алгоритма адаптивной интерполяции для моделирования динамических систем с интервальными параметрами в соответствии с работами [23, 26, 27].

Выполним описание алгоритма на примере рассматриваемой выше системы ОДУ:

$$\begin{cases} \frac{dx(t)}{dt} = f(x(t), V(t), R_{on}, R_{off}, V_n, V_p, \mu_v, \delta), \\ x(t_0) = 0, \\ (R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)^T \in \mathcal{X}, \\ t \in [t_0, t_N], \end{cases}$$

где f – функция задающая скорость изменения состояния мемристора. Решение данной системы в каждый момент времени t_k можно рассматривать как функцию от точечных значений интервальных параметров $x_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$. Цель алгоритма – построить функцию $P_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$, интерполирующую $x_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$ с контролируемой точностью. Дальнейшее определение интервальной оценки тока (фактически множества \mathbf{I}_k)

$$\left[\underline{I}_k, \overline{I}_k \right] = \left[\min_{(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)^T \in \mathcal{X}} I_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta), \max_{(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)^T \in \mathcal{X}} I_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) \right]$$

сводится к двум задачам условной оптимизации для явной функции, которые можно решить, например, с помощью методов, представленных в работе [34].

В начальный момент времени t_0 функция P_0 определяется тривиальным образом: $P_0(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) = x(t_0) = 0$. Построим с помощью индукции решение для произвольного момента времени. Пусть в момент времени t_k известна $P_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$. Получение $P_{k+1}(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta)$ сводится к интерполяции неявной функции $z(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta) = \hat{x}(t_{k+1})$, заданной в виде системы ОДУ:

$$\begin{cases} \frac{d\hat{x}(t)}{dt} = f(\hat{x}(t), V(t), R_{on}, R_{off}, V_n, V_p, \mu_v, \delta), \\ \hat{x}(t_k) = P_k(R_{on}, R_{off}, V_n, V_p, \mu_v, \delta), t \in [t_k, t_{k+1}]. \end{cases}$$

Традиционно интерполяционный полином P_k строится по определенному набору узлов, которые образуют сетку. Поэтому сначала выполняется перенос решений, которые соответствуют узлам, на $k+1$ временной слой, а далее в зависимости от значения погрешности интерполяции происходит адаптация.

Далее рассмотрим модификацию модели (2.3):

$$\frac{dx(t)}{dt} = a \cdot V(t)^s \begin{cases} 1 - (1-x)^{2\text{round}\left(\frac{b}{|V(t)+c}\right)}, & V(t) \leq -V_{thr}, \\ 1 - x^{2\text{round}\left(\frac{b}{|V(t)+c}\right)}, & V(t) > V_{thr}, \\ 0, & -V_{thr} < V(t) \leq V_{thr}, \end{cases}$$

$$I(t) = x(t)^n \beta \sinh(\alpha_M V(t)) + \chi [\exp(\gamma V(t)) - 1] + \delta, \quad (2.7)$$

$$R(t) = \frac{V(t)}{I(t)},$$

где x – переменная состояния; $a \in [1.000, 1.119]$ В⁻⁵ – постоянная, определяемая свойствами материала; $s=5$ – нечетное целое число; $I(t)$, $V(t)$, $R(t)$ – текущие значения тока, напряжения и сопротивления; $V_{thr} = 0.4$ В – пороговое значение напряжения активации; $n=9$, $\beta = 9 \cdot 10^{-5}$ В, $\chi = 1.5 \cdot 10^{-4}$ В, $\alpha_M \in [1.674, 2.114]$ В⁻¹, $\gamma \in [-0.019, 0.415]$ В⁻¹, $\delta \in [-0.011, 1.464]$ мкА – подгоночные параметры в выражении для тока; round – функция получения целочисленного результата; $b=15$ В, $c=2$ В – подгоночные коэффициенты основного уравнения. Данные значения параметров получены с помощью подхода описанного в работе [31] для наилучшего воспроизведения экспериментальных данных по мемристору на основе оксида гафния [16]. На рис. 2.11 показана зависимость напряжения от времени $V(t)$, на рис. 2.12 черной кривой представлена экспериментальная вольтамперная характеристика (ВАХ) и на рис. 2.13 восстановленная по $V(t)$ и ВАХ зависимость тока от времени. Серым цветом показаны полученные модельные интервальные оценки.

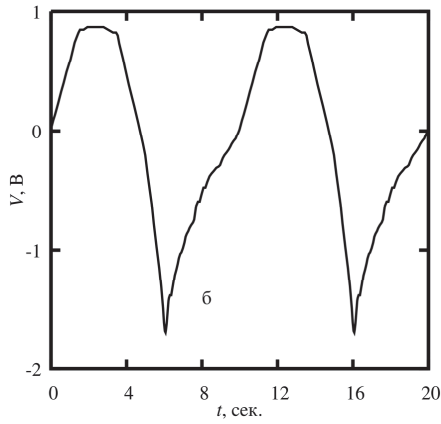


Рис. 2.11. Исходная зависимость напряжения от времени в эксперименте

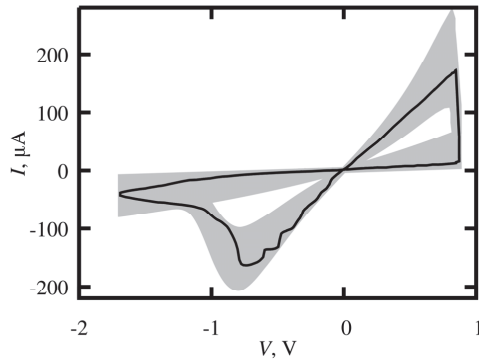


Рис. 2.12. Экспериментальная и модельная вольтамперная характеристика

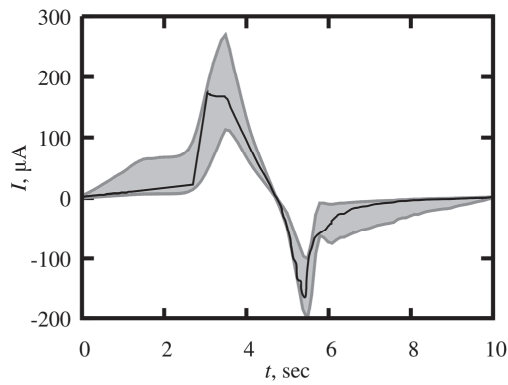


Рис. 2.13. Экспериментальная и модельная зависимость тока от времени

Обе приведенные модели с интервальными параметрами дают интервальные оценки, которые полностью содержат в себе экспериментальные данные, что дополнительно подтверждает эффективность подхода к интервальной параметрической идентификации динамических систем приведенного в работе [31].

ГЛАВА 3

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ АНАЛОГОВОЙ САМООБУЧАЮЩЕЙСЯ ИМПУЛЬСНОЙ НЕЙРОННОЙ СЕТИ

Вопрос ускорения вычислений является актуальным в любое время. Сначала производители увеличивали частоту работы процессора, потом увеличивали количество ядер в одном процессоре, далее адаптировали графические процессоры для вычислений общего характера, а в настоящее время перспективным представляется переход от цифровых вычислений к аналоговым вычислениям, так как они выполняются на порядки быстрее [35]. При разработке аналоговых вычислительных устройств часто используется относительно новый электрический элемент – мемристор, представляющий собой резистор, проводимость которого зависит от суммарного протекшего через него электрического заряда.

Объединение мемристоров в матрицу называется мемристорным кроссбаром. Благодаря закону Ома и закону Кирхгофа, на основе кроссбара можно выполнять быстрое аналоговое произведение матрицы на вектор [36, 37]. Помимо эффективного применения мемристивных элементов в рамках параллелизации матрично-векторных операций, перспективным представляется их использование для аналоговой реализации самообучающихся импульсных нейронных сетей (ИНС).

Импульсные нейронные сети являются третьим поколением нейронных сетей. В данном виде сетей обмен информацией осуществляется в виде импульсов, что в наибольшей степени соответствует физиологии биологического мозга. Обучение ИНС основывается на локальных правилах изменения весов. Существуют аппаратные реализации ИНС на полупроводниковых элементах, в частности, проект TrueNorth [38]. Мемристивные элементы используются, не только в ИНС, но и в глубоких нейронных сетях [39], однако зачастую в этом случае задействуются

алгоритмы обратного распространения ошибок, обладающие нелокальным характером и большими вычислительными сложностями.

Существует ряд работ, посвященных вопросам реализации ИНС на основе мемристоров. В частности, работа [40] направлена на изучение экспериментальных и теоретических подходов к поиску эффективных правил обучения. В [41] представлен подход к моделированию нейронных сетей, основанный на реализации металлооксидных гетероструктур с энергонезависимой памятью и многоуровневым резистивным переключением. В работе [42] экспериментально демонстрируется протокол обучения, который является не чувствительным к начальному состоянию мемристоров, а так же к их различиям в рамках одной сети.

Обучение ИНС на аппаратном уровне традиционно основывается на правиле Хебба и синаптической пластичности. Используется метод Spike Timing Dependent Plasticity (STDP), согласно которому изменение весов синапсов нейрона зависит от разницы во времени между входным и выходным импульсом [43–48]. Те синаптические связи, которые вызвали активацию нейрона, усиливаются, а другие ослабляются. Со схемотехнической позиции метод STDP реализуется за счет 1T1R-кроссбар-архитектуры, в которой каждому мемристорному элементу соответствует один транзистор, отвечающий за изменение проводимости, и наличия обратной связи у каждого нейрона со всеми синапсами.

В соответствии с тем как рассматривались модели мемристоров (детерминированная, стохастическая и интервальная) в последующих разделах будут описаны имитационные модели сетей на их основе.

3.1. Детерминированная модель нейронной сети

Вопросы построения и моделирования схемотехнических решений для нейроморфных сетей рассматривались в [49–53]. Для таких сетей показана способность SDTP-обучения к распознаванию образов на аналоговом (аппаратном) уровне. В настоящем разделе сформулирована строгая

детерминированная математическая модель нейроморфной сети в виде нелинейной динамической системы.

Рассмотрим работу схемотехнического решения однослойной самообучающейся аналоговой импульсной нейронной сети с мемристивными элементами в качестве синаптических весов (рис. 3.1). Пришедшие на вход импульсы V_g открывают соответствующие транзисторы, что приводит к протеканию токов через мемристоры с последующим их суммированием в нейронах. Схемотехническая модель нейрона (серый кружок на рис. 3.1) представляет собой параллельную RC цепь и абстрактный генератор импульсов G (рис. 3.2). Как только значение потенциала на конденсаторе превышает некоторый порог, его потенциал сбрасывается, и генератор импульсов выдает выходной сигнал V_{out} и сигнал обратной связи V_{te} . Помимо этого, в обратной связи постоянно поддерживается некоторый небольшой потенциал, необходимый для функционирования сети в обычном режиме. Значения емкости C_{int} и сопротивления R_{int} отвечают за скорость накопления и снижения потенциала в нейроне.

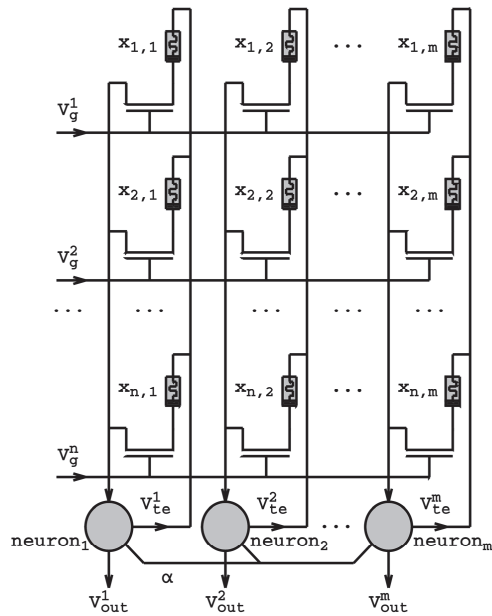


Рис. 3.1. Схемотехническая реализация импульсной нейронной сети

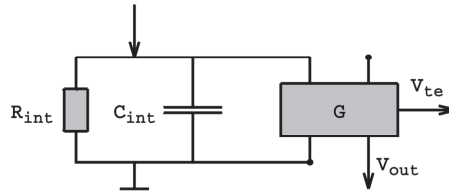


Рис. 3.2. Схемотехническая реализация нейрона

Процесс обучения сети происходит согласно правилу STDP (те синаптические связи, которые привели к активации нейрона, усиливаются, другие – ослабевают). Данный механизм обучения реализуется путем обратной связи у нейронов (V_{te}). В момент активации нейрона по каналу обратной связи с задержками поступают два противоположных по знаку импульса. Если на синапсе есть активность и пришел положительный импульс обратной связи, то значение проводимости соответствующего мемристора увеличивается, а если пришел отрицательный импульс обратной связи, то проводимость мемристора уменьшается. На рис. 3.3 проиллюстрирован процесс изменения синаптических весов на примере одного синапса.

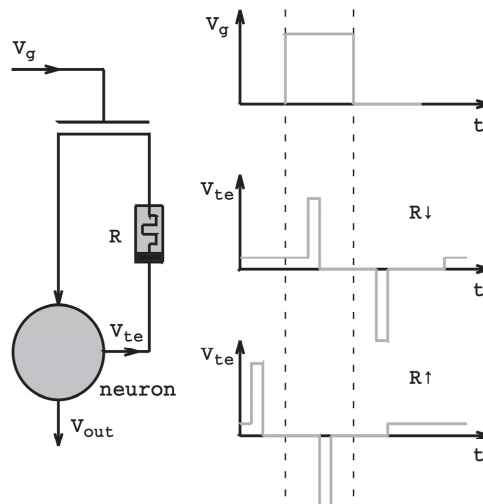


Рис. 3.3. Схемотехническая реализация правила обучения STDP

Обучение сети происходит следующим образом: на вход в сеть с равной вероятностью подается или произвольный шум, или один из заранее определенных шаблонов. Спустя некоторое время сеть адаптируется к распознаванию шаблонов. Подача в сеть шума необходима для уменьшения проводимости мемристоров, которые не соответствуют никакому шаблону. С вероятностью $1/2$ после того как нейрон активировался в сеть будет подан шум, и как раз отрицательные импульсы в обратной связи окажут воздействия на соответствующие мемристоры, уменьшив их проводимость.

В случае нескольких шаблонов в выходном слое будет несколько нейронов с дополнительной связью (на рис. 3.1 помечена буквой α). Когда происходит активация одного нейрона, то он подавляет остальные нейроны (уменьшает значение их потенциала). Распределение шаблонов по нейронам происходит в процессе обучения.

Сформулируем комплексную математическую модель однослойной самообучающейся импульсной нейронной сети (рис. 3.1). Определим основные переменные модели:

n – количество входов, m – количество нейронов;

V_g^i – текущее значение напряжения на i -м входе нейронной сети;

V_{te}^j – текущее значение напряжения в обратной связи j -го нейрона;

V_{out}^j – текущее значение напряжения на выходе j -го нейрона;

τ_j – время, прошедшее после последней активации j -го нейрона;

V_{int}^j – напряжение на конденсаторе j -го нейрона;

R_{int}, C_{int} – значение сопротивления и емкости у нейронов;

$V_{te}^+, V_{te}^-, V_{te}^0$ – значения амплитуды импульсов обратной связи и значение напряжения по умолчанию;

V_{out}^+ – амплитуда выходного импульса;

V_{th} – уровень напряжения активации нейрона;

$R_{i,j}$ – значение сопротивления мемристора i -го синапса j -го нейрона;

$x_{i,j}$ – состояние мемристора i -го синапса j -го нейрона, $x_{i,j} \in [0, 1]$;
 τ_r – длительность сигнала в обратной связи после активации нейрона;
 τ_s – длительность одного импульса в сигнале обратной связи, $2\tau_s < \tau_r$;
 τ_{out} – длительность одного импульса на выходе сети;
 α – коэффициент подавления.

Математическая модель задается следующими соотношениями:

$$\frac{dx_{i,j}}{dt} = \begin{cases} F_X \left(\frac{V_{te}^j - V_{int}^j}{R_{i,j}}, V_{te}^j - V_{int}^j, x_{i,j} \right), & V_g^i(t) > 0, \\ 0, & V_g^i(t) = 0, \end{cases} \quad (3.1)$$

$$R_{i,j} = F_R(V_{te}^j - V_{int}^j, x_{i,j}),$$

$$\frac{dV_{int}^j}{dt} = \frac{1}{C_{int}} \left[\sum_{i=1}^n \frac{\hat{V}_{te}^j - V_{int}^j}{R_{i,j}} - \frac{V_{int}^j}{R_{int}} \right] - \max_{i=1,m} [\theta(V_{int}^i - V_{th}) \hat{\alpha}_{i,j}] \delta \left(\prod_{i=1}^m (V_{int}^i - V_{th}) \right) V_{int}^j, \quad (3.2)$$

$$\frac{d\tau_j}{dt} = 1 - \delta(V_{int}^j - V_{th}) \tau_j, \quad (3.3)$$

$$V_{te}^j = \begin{cases} V_{te}^+, & \tau_j \leq \tau_s, \\ 0, & \tau_s < \tau_j \leq \frac{\tau_r}{2}, \\ V_{te}^-, & \frac{\tau_r}{2} < \tau_j \leq \frac{\tau_r}{2} + \tau_s, \\ 0, & \frac{\tau_r}{2} + \tau_s < \tau_j \leq \tau_r, \\ V_{te}^0, & \tau_r < \tau_j, \end{cases} \quad (3.4)$$

$$V_{out}^j = \begin{cases} V_{out}^+, & \tau_j \leq \tau_{out}, \\ 0, & \tau_{out} < \tau_j, \end{cases} \quad (3.5)$$

$$\hat{\alpha}_{i,j} = 1 - \alpha(1 - \delta_{ij}), \quad i = \overline{1, n}, \quad j = \overline{1, m},$$

– и дополняется начальными условиями

$$x_{i,j}(0) = \text{rand}[0, 1], \quad V_{int}^j(0) = 0, \quad \tau_j(0) > \max(\tau_r, \tau_{out}),$$

где $\hat{V}_{te}^j = \max(0, \min(V_{te}^j, V_{te}^0))$ и отвечает за то, чтобы импульсы обратной связи не вносили вклад в накопление потенциала внутри нейрона; δ_{ij} – символ

Кroneckera; $\delta(x)$ – дельта-функция; $\theta(x)$ – функция Хэвисайда. Соотношения (3.1) задают модель мемристора. Функция $F_x(I, v, x)$ определяет скорость изменения переменной состояния в зависимости от тока (I), напряжения (v) и текущего состояния (x). Функция $F_R(x, v)$ определяет зависимость сопротивления мемристора от состояния и приложенного напряжения.

Соотношение (3.2) задает модель нейрона, который представляет собой параллельную RC цепочку (рис. 3.2), включенную последовательно с резистором (все мемристоры синапсов на уровне нейрона можно рассматривать как один резистор).

Уравнение (3.3) реализует механизм счетчика времени после последней активации нейрона. Как только напряжение V_{int}^j на конденсаторе достигает порогового значения V_{th} , переменная τ_j обнуляется. Аналогично происходит в уравнении (3.3): после активации нейрона накопленный им потенциал сбрасывается, а у других нейронов – уменьшается прямо пропорционально коэффициенту α .

Уравнения (3.4) и (3.5) определяют форму импульсов в обратной связи и на выходе нейронной сети. В начальный момент времени переменная τ_j выбирается таким образом, чтобы избежать преждевременного появления импульсов в обратной связи и на выходе.

Уравнения (3.3) – (3.5)

и член
$$\max_{i=1, m} \left[\theta(V_{int}^i - V_{th}) (1 - \alpha (1 - \delta_{ij})) \right] \delta \left(\prod_{i=1}^m (V_{int}^i - V_{th}) \right) V_{int}^j$$

в уравнении (3.2) описывают логику абстрактного генератора импульсов G (рис. 3.2). В момент, когда напряжение V_{int}^i в i -м нейроне достигает порогового значения V_{th} , V_{int}^i сбрасывается, а напряжения остальных нейронов уменьшаются в α раз (для этого используются дельта-функция и функция Хэвисайда).

Отметим, что уравнения (3.2) (нейронная модель) не связаны напрямую с дифференциальной частью уравнений (3.1) (мемристорная модель). Это

позволяет реализовать различные типы моделей динамики переключения мемристоров.

Приведенная математическая модель описывает только один слой нейронной сети. Для моделирования многослойной сети достаточно соединить выход k -го слоя со входом $(k+1)$ -го слоя: $V_g^{i(k+1)}(t) = V_{out}^{j(k)}(t), i = j$. Таким образом, область применения описанной математической модели не ограничивается однослойной сетью.

При моделировании работы схемотехнического решения в качестве модели мемристора сначала используется модель (2.3). Соответственно, функции $F_X(I, V, x)$ и $F_R(V, x)$, входящие в систему уравнений (3.1) – (3.5), определяются следующим образом:

$$F_X(I, V, x) = a \cdot V^s \begin{cases} 1 - (1-x)^{2 \operatorname{round}\left(\frac{b}{|V|+c}\right)}, & V \leq -V_{thr}, \\ 1 - x^{2 \operatorname{round}\left(\frac{b}{|V|+c}\right)}, & V > V_{thr}, \\ 0, & -V_{thr} < V \leq V_{thr}, \end{cases}$$

$$F_R(V, x) = \frac{V}{x^n \beta \sinh(\alpha_M V) + \chi [\exp(\gamma V) - 1]}.$$

Рассмотрим сеть, состоящую из одного нейрона ($m = 1$, поэтому индекс j в дальнейшем опускается) с 64 синапсами (рис. 3.4).

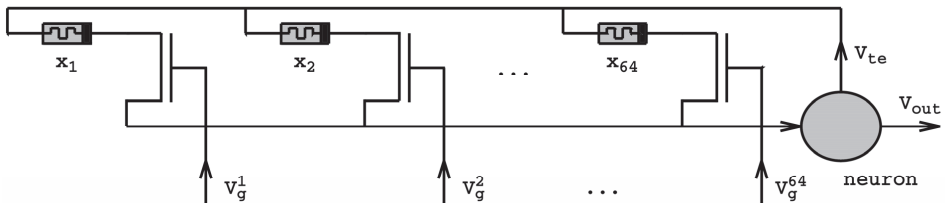


Рис. 3.4. Нейронная сеть, состоящая из одного нейрона

Динамическая система имеет следующий вид:

$$\left\{ \begin{array}{l} \frac{dx_i}{dt} = \begin{cases} F_X \left(\frac{V_{te}(\tau) - V_{int}}{R_i}, V_{te}(\tau) - V_{int}, x_i \right), V_g^i(t) > 0, \\ 0, V_g^i(t) = 0, \end{cases} \\ \frac{dV_{int}}{dt} = \frac{1}{C_{int}} \left[\sum_{\substack{i=1 \\ V_g^i(t) > 0}}^n \frac{\hat{V}_{te}(\tau) - V_{int}}{R_i} - \frac{V_{int}}{R_{int}} \right] - \delta(V_{int} - V_{th})V_{int}, \\ \frac{d\tau}{dt} = 1 - \delta(V_{int} - V_{th})\tau, \\ R_i = F_R(x_i, V_{te}(\tau) - V_{int}), \\ x_i(0) = \text{rand}[0, 1], V_{int}(0) = 0, \tau(0) > \max(\tau_r, \tau_{out}), \\ i = \overline{1, n}. \end{array} \right.$$

и дополняется соотношениями (3.4) и (3.5).

Значения параметров: $n = 64$, $R_{int} = 1 \text{ кОм}$, $C_{int} = 45 \text{ мкФ}$, $V_{te}^+ = 1.5 \text{ В}$, $V_{te}^- = -1.6 \text{ В}$, $V_{te}^0 = 10 \text{ мВ}$, $V_{out}^+ = 2 \text{ В}$, $V_{th} = 3 \text{ мВ}$, $\tau_r = 20 \text{ мс}$, $\tau_s = 2 \text{ мс}$, $\tau_{out} = 10 \text{ мс}$. Значения R_{int} , C_{int} подобраны так, чтобы коэффициент релаксации ($= R \cdot C$) был равен 45 мс, что соответствует значению, приведенному в работе [44].

В процессе моделирования каждые 10 мс компоненты вектора $V_g(t)$ могут с равной вероятностью или быть случайным шумом (V_g^i имеет дискретное распределение (табл. 1)) или принимать определенные значения, которые задаются в соответствии с распознаваемым шаблоном (рис. 3.5). Для наглядности вектор $V_g(t)$ здесь записывается в виде матрицы:

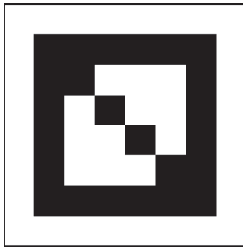


Рис. 3.5. Распознаваемый шаблон

$$V_g = \begin{pmatrix} 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\ 0, 2, 2, 2, 2, 2, 2, 2, 0, \\ 0, 2, 2, 0, 0, 0, 2, 0, \\ 0, 2, 0, 2, 0, 0, 2, 0, \\ 0, 2, 0, 0, 2, 0, 2, 0, \\ 0, 2, 0, 0, 0, 2, 2, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 0, 0, 0, 0, 0, 0, 0 \end{pmatrix} \text{ В}$$

V_g^i	0 В	2 В
P	0.85	0.15

Табл. 1. Дискретное распределение шума для обучения сети с одним нейроном

Далее приведены графики поведения напряжений нейронной сети в зависимости от времени (рис. 3.6, 3.7 и 3.8).

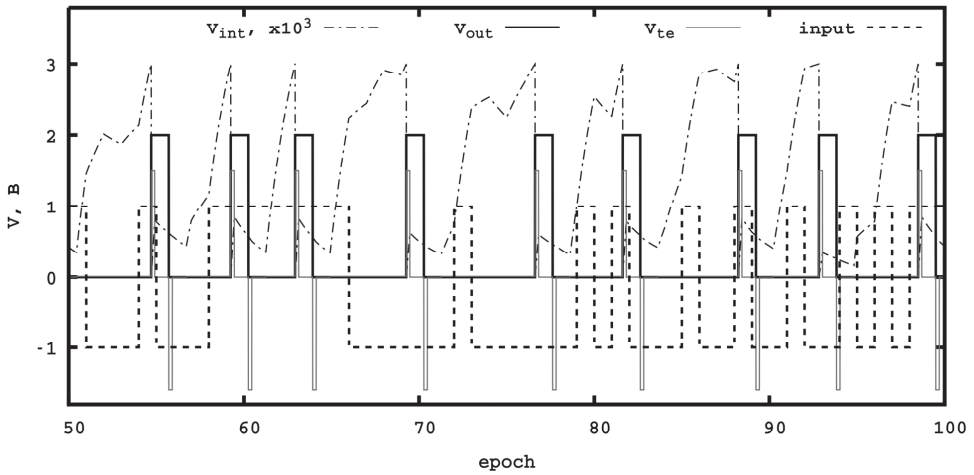


Рис. 3.6. Процесс обучения нейронной сети с 50-ой по 100-ю эпоху

Одна эпоха обучения равна 10 мс, то есть тому времени, на протяжении которого нейронной сети демонстрируется образец или случайный шум. Узким пикам соответствует напряжение в обратной связи нейрона. Кривая линия отображает накопленный потенциал в конденсаторе нейрона. Как только его значение достигает порогового значения (3 мВ, на графике с учетом масштаба 3 В), происходит генерация выходного импульса и импульсов обратной связи. Пунктирная линия (input) показывает, что конкретно в каждый момент времени демонстрируется нейронной сети. Значению -1 соответствует шум, а значению 1 соответствует образец.

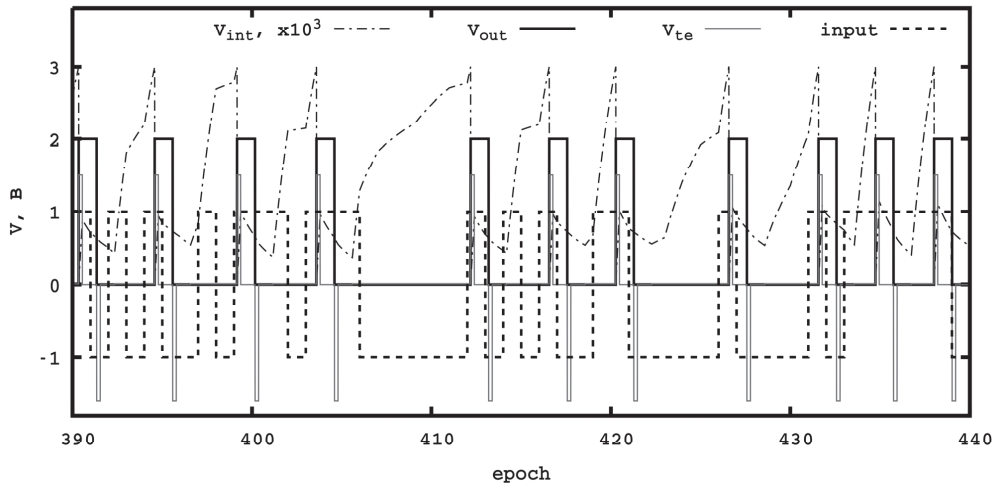


Рис. 3.7. Процесс обучения нейронной сети с 390-й по 440-ю эпоху

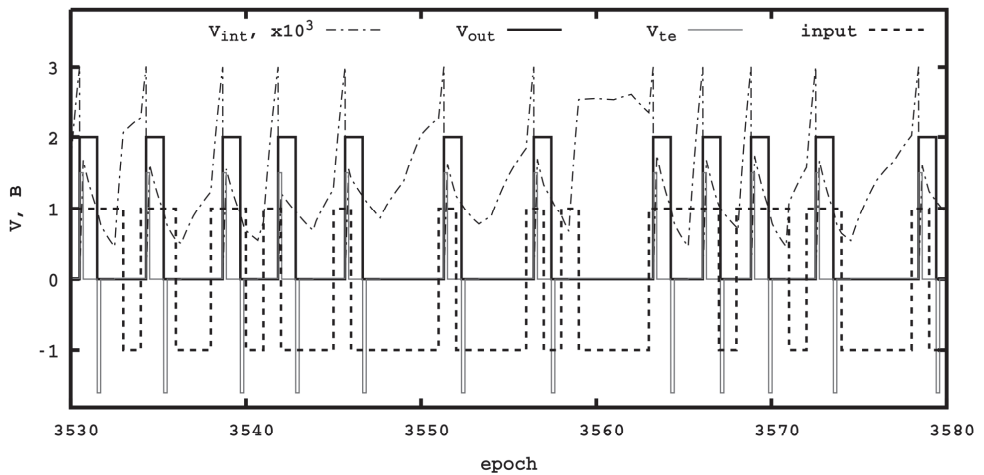


Рис. 3.8. Работа нейронной сети в конце процесса обучения

Если в начале процесса обучения (рис. 3.6) наблюдаются многочисленные пропуски шаблона и ложное срабатывание на шум (четвертая и пятая активация нейрона), то спустя некоторое количество эпох (рис. 3.7) сеть уже практически всегда реагирует на шаблон и игнорирует шум. В конце процесса обучения (рис. 3.8) сеть не срабатывает на шаблон только в том случае, если до этого была активация нейрона и импульсы

обучения не дают возможность работать сети в обычном режиме. Отметим, что в моменты демонстрации шума сети потенциал внутри нейрона не только возрастает, но и убывает.

На рис. 3.9 показан процесс адаптации синаптических весов к распознаваемому образцу. Цвет соответствует значению сопротивления соответствующего мемристора: чем темнее, тем проводимость больше; чем светлее, тем меньше. В начальный момент времени все веса инициализируются случайными значениями и в процессе работы сети постепенно изменяются. Примерно с 750-ой эпохи начинает просматриваться шаблон, распознаванию которого обучается сеть: происходит запоминание информации нейронной сетью.

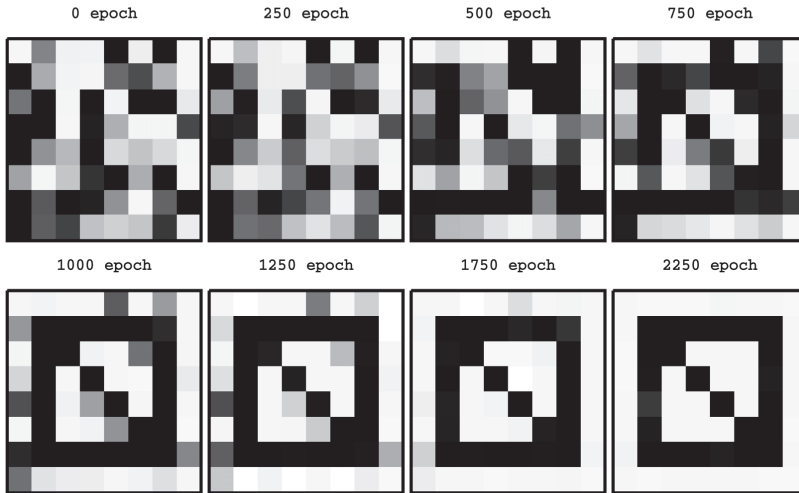


Рис. 3.9. Изменение синаптических весов в процессе обучения

Далее рассматривается задача распознавания двух образцов. На рис. 3.10 приведена схемотехническая реализация сети. В отличие от предыдущего примера, здесь происходит взаимодействие между нейронами. При активации одного нейрона значение потенциала другого нейрона уменьшится прямо пропорционально коэффициенту подавления $\alpha = 0.4$. Порог активации $V_{th} = 4$ мВ, а остальные параметры модели прежние.

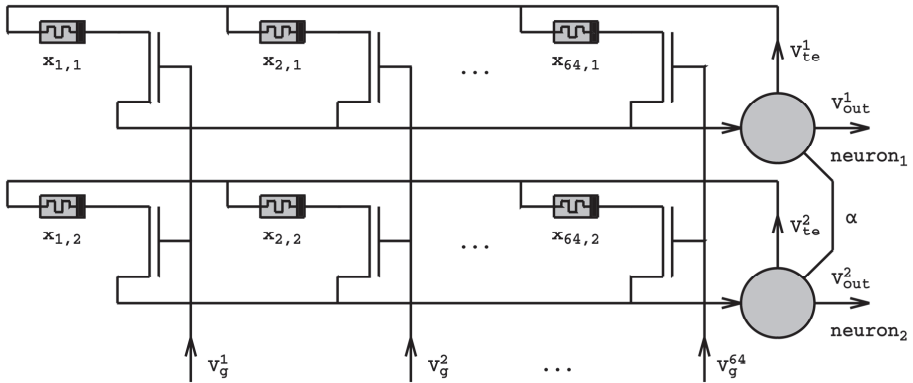


Рис. 3.10. Нейронная сеть, состоящая из двух нейронов

Как и ранее, каждые 10 мс компоненты вектора $V_g(t)$ могут с равной вероятностью быть случайным шумом (V_g^i имеет дискретное распределение (табл. 2)) или принимать с равной вероятностью одно из двух значений, которые определяются распознаваемыми шаблонами (буква «а» и квадрат):

$$V_g^{(1)} = \begin{pmatrix} 0, 0, 0, 0, 0, 0, 0, 0 \\ 0, 0, 2, 2, 2, 2, 0, 0 \\ 0, 2, 2, 0, 0, 2, 2, 0 \\ 0, 2, 2, 0, 0, 2, 2, 0 \\ 0, 2, 2, 2, 2, 2, 2, 0 \\ 0, 2, 2, 2, 2, 2, 2, 0 \\ 0, 2, 2, 2, 2, 2, 2, 0 \\ 0, 2, 2, 0, 0, 2, 2, 0 \\ 0, 0, 0, 0, 0, 0, 0, 0 \end{pmatrix}^T, \quad B, V_g^{(2)} = \begin{pmatrix} 2, 2, 2, 2, 2, 2, 2, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 0, 0, 0, 0, 0, 0, 2 \\ 2, 2, 2, 2, 2, 2, 2, 2 \end{pmatrix}^T \quad B \quad (3.6)$$

V_g^i	0 В	2 В
P	0.8	0.2

Табл. 2. Дискретное распределение шума для обучения сети с двумя нейронами

На рис. 3.11 и 3.12 показаны графики напряжений для каждого нейрона в отдельности. Пунктирная линия (input) соответствует входу нейронной

сети: значение 0 – шум, значение -1 – первый образец, значение 1 – второй образец. На графиках потенциала конденсаторов четко видны резкие скачки, это связано с активацией одного из нейронов и подавлением другого нейрона.

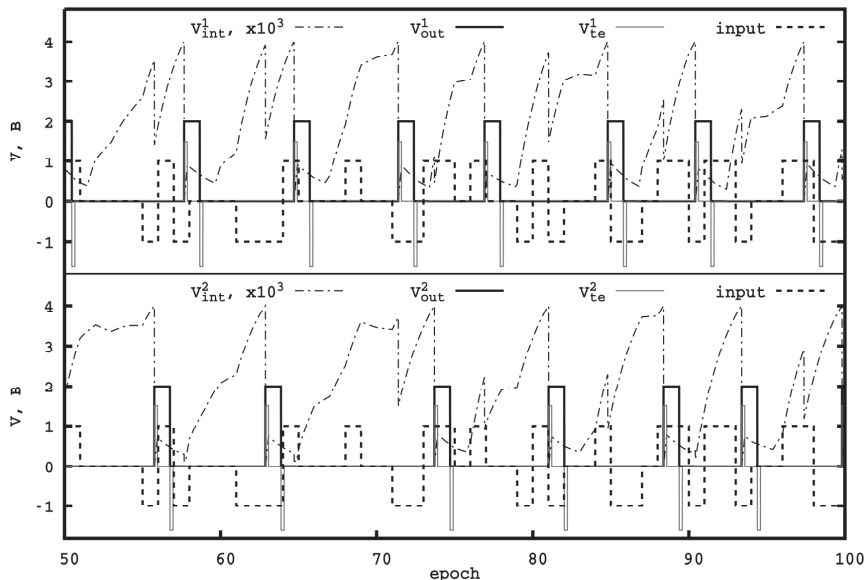


Рис. 3.11. Процесс обучения нейронной сети с 50-ой по 100-ю эпоху

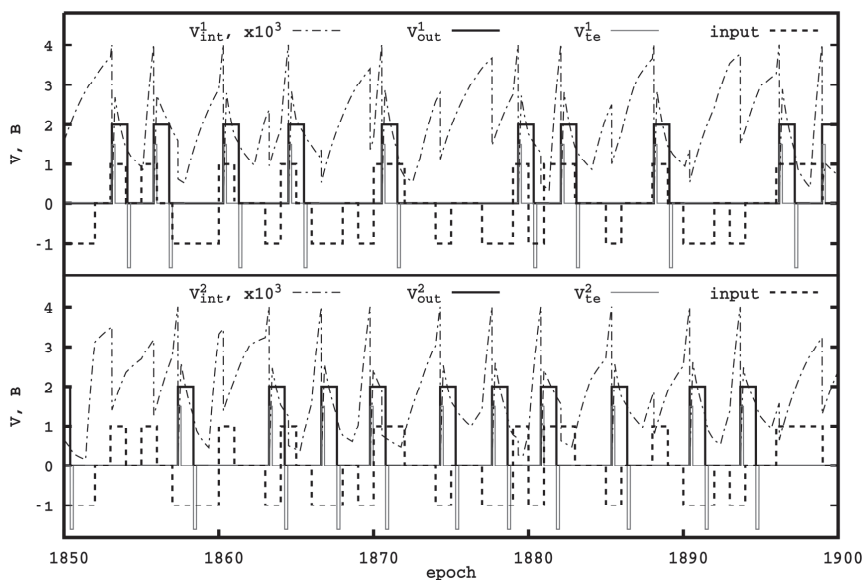


Рис. 3.12. Работа нейронной сети в конце процесса обучения

С точки зрения качества распознавания, в начале процесса обучения (рис. 3.11) проявляется хаос: один и тот же нейрон активируется на разные образцы либо вообще оба нейрона игнорируют шаблоны. В конце процесса обучения (рис. 3.12) каждый нейрон срабатывает только на свой определенный входной сигнал.

На рис. 3.13 показан процесс адаптации синаптических весов к распознаваемым образцам. В отличие от предыдущего примера, здесь шаблоны начинают просматриваться примерно с 600-ой эпохи.

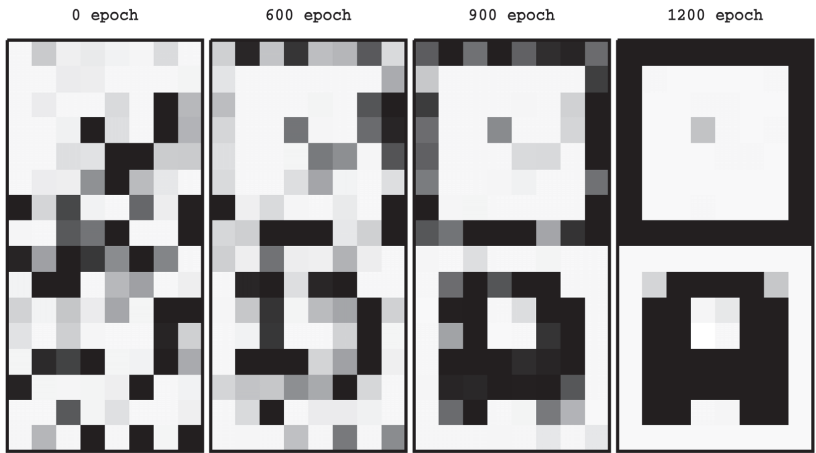


Рис. 3.13. Изменение синаптических весов в процессе обучения

Отметим, что результаты, полученные с использованием разработанной имитационной модели, хорошо согласуются с результатами, опубликованными в работах [44, 49–53].

Далее рассмотрим моделирование процесса функционирования нейроморфной сети с мемристивными элементами в качестве синаптических весов основанных на разных оксидах на примере распознавания пяти шаблонов шаблонов (рис. 3.14) $m = 5$ и $n = 8 \times 8 = 64$. В процессе моделирования работы нейронной сети, каждую эпоху обучения (равную $\tau_r / 2$ сек), компоненты вектора $V_g(t)$ могут с равной вероятностью или быть случайным шумом (V_g^i имеет дискретное распределение), или принимать

с равной вероятностью одно из пяти значений, которые задаются в соответствии с распознаваемыми шаблонами. Запишем вектор V_g (для наглядности в виде матрицы) для первых двух шаблонов:

$$V_g^{(1)} = \begin{pmatrix} 0, 0, 0, 0, 0, 0, 0, 0, \\ 0, 0, 2, 2, 2, 2, 0, 0, \\ 0, 2, 2, 0, 0, 2, 2, 0, \\ 0, 2, 2, 0, 0, 2, 2, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 2, 2, 0, 0, 2, 2, 0, \\ 0, 0, 0, 0, 0, 0, 0, 0 \end{pmatrix}^T, \quad V_g^{(2)} = \begin{pmatrix} 0, 0, 0, 0, 2, 2, 0, 0, \\ 0, 2, 2, 2, 2, 0, 0, 0, \\ 0, 2, 0, 0, 2, 0, 0, 2, \\ 0, 2, 0, 0, 2, 2, 2, 2, \\ 2, 2, 2, 2, 0, 0, 2, 0, \\ 2, 0, 0, 2, 0, 0, 2, 0, \\ 0, 0, 0, 2, 2, 2, 2, 0, \\ 0, 0, 2, 2, 0, 0, 0, 0 \end{pmatrix}^T \quad B.$$

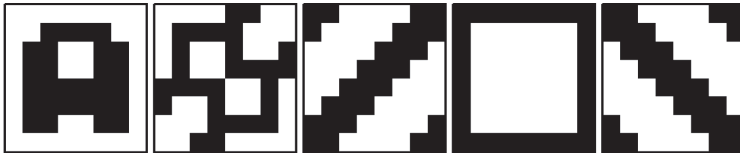


Рис. 3.14. Распознаваемые шаблоны

В зависимости от используемой модели мемристора выполняется настройка параметров математической модели (3.1) – (3.5). Для модели (2.1), которая соответствует мемристору на основе оксида титана (TiO_2), имеем следующие значения параметров: $R_{int} = 200$ Ом, $C_{int} = 45$ мкФ, $V_{te}^+ = 0.7$ В, $V_{te}^- = -0.9$ В, $V_{te}^0 = 10$ мВ, $V_{out}^+ = 2$ В, $V_{th} = 9$ мВ, $\tau_r = 3$ мс, $\tau_s = 50$ мкс, $\tau_{out} = 1.5$ мс. Значения V_{te}^+ и V_{te}^- подбираются из расчета, чтобы инициировать переключение мемристора, и чем они больше по абсолютному значению, тем переключение мемристора происходит быстрее. Длительность импульсов τ_s в обратной связи также влияет на скорость переключения мемристора: чем больше длительность, тем быстрее произойдет переключение. Значение порога активации нейрона V_{th} не должно превышать напряжение по умолчанию в обратной связи V_{te}^0 , иначе нейрон не активируется. В то же время напряжение по умолчанию V_{te}^0 не должно быть слишком большим,

чтобы не вызывать переходные процессы в мемристорах. R_{int} и C_{int} подбираются в соответствии с диапазоном изменения сопротивления мемристивных элементов и отвечают за быстроту накопления заряда в нейроне, поэтому дополнительно должны быть согласованы с длительностью одной эпохи. Параметры τ_{out} и V_{out}^+ в случае однослойной сети не оказывают никакого влияния.

V_g^i имеет дискретное распределение: $V_g^i = 0\text{В}$ с вероятностью 0.73 и $V_g^i = 2\text{В}$ с вероятностью 0.27. На рис. 3.15 показан процесс адаптации синаптических весов к распознаваемым образцам. Цвет соответствует значению переменной состояния соответствующего мемристора: чем темнее, тем проводимость больше; чем светлее, тем меньше. В начальный момент времени все веса инициализируются случайными значениями, и в процессе работы сети постепенно изменяются. Примерно с 700-ой эпохи начинают просматриваться шаблоны, распознаванию которых обучается сеть: происходит запоминание информации нейронной сетью.

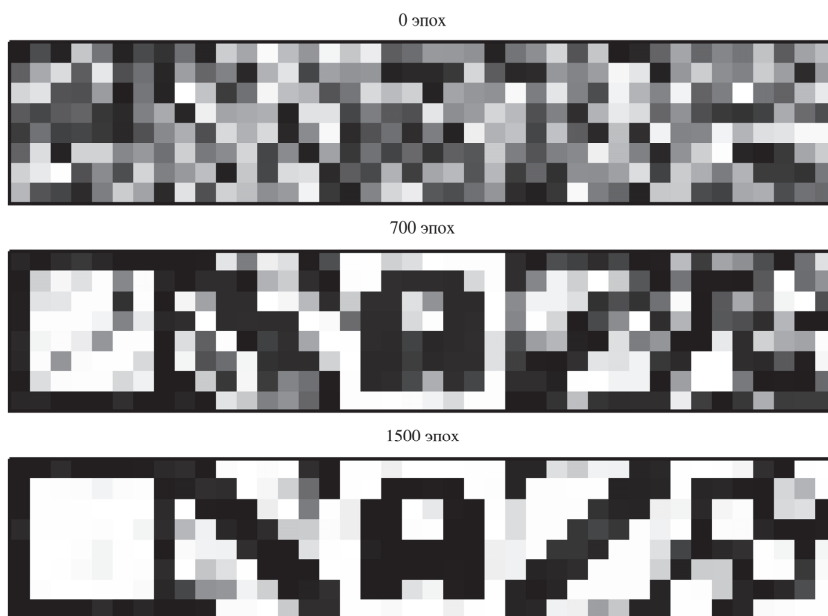


Рис. 3.15. Изменение синаптических весов в процессе обучения нейроморфной сети на основе оксида титана

Далее рассматривается модель мемристора (2.3), которая соответствует оксиду гафния (HfO_2). Параметры математической модели нейронной сети: $R_{int} = 1$ кОм, $C_{int} = 45$ мкФ, $V_{te}^+ = 1.55$ В, $V_{te}^- = -1.6$ В, $V_{te}^0 = 10$ мВ, $V_{out}^+ = 2$ В, $V_{th} = 2.5$ мВ, $\tau_r = 15$ мс, $\tau_s = 1$ мс, $\tau_{out} = 7.5$ мс. Настройка параметров осуществлялась по аналогии с моделью мемристора на основе оксида титана. $V_g^i = 0$ В с вероятностью 0.85 и $V_g^i = 2$ В с вероятностью 0.15. На рис. 3.16 показано изменение синаптических весов сети в процессе обучения.

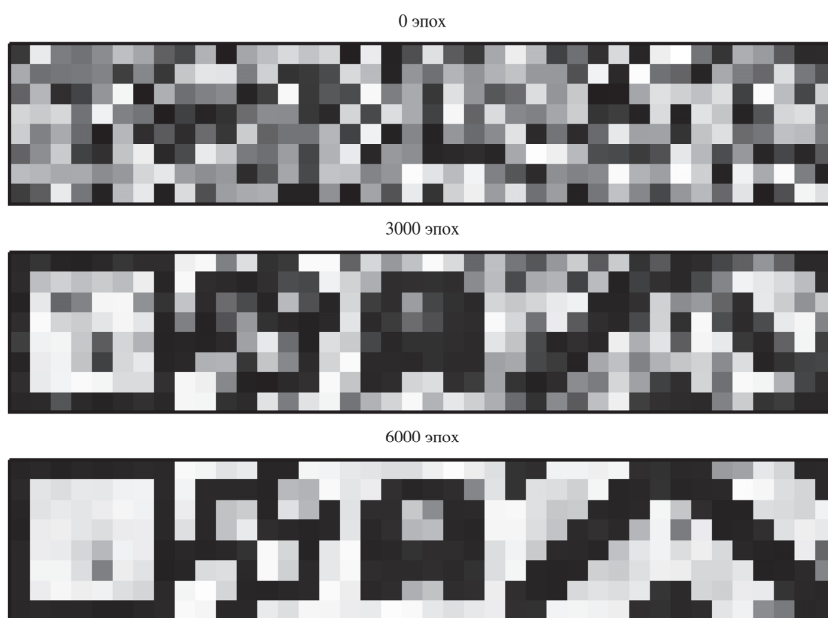


Рис. 3.16. Изменение синаптических весов в процессе обучения нейроморфной сети на основе оксида гафния

В отличие от предыдущего примера здесь сеть обучалась дольше. Отметим, что приведенные наборы параметров нейронной сети являются одним из возможных наборов и не являются оптимальными с позиции скорости обучения сети. В процессе обучения шаблоны распределяются по нейронам произвольным образом, поэтому в двух приведенных примерах нейроны, отвечающие за один и тот же шаблон, разные.

3.2. Стохастическая модель нейронной сети

В данном разделе исследуется влияние стохастических свойств мемристоров на работу нейроморфной сети. Практический интерес представляет вопрос использования значений параметров сети, определенных с помощью детерминированной модели, в условиях стохастической динамики мемристивных элементов. Другими словами, возможно ли выполнять настройку сети только на детерминированной модели. Кроме этого, важным представляется понимание того, как уровень шума будет влиять на скорость обучения и при каком уровне аддитивного шума нейронная сеть не сможет нормально функционировать.

Учет стохастических свойств мемристивных элементов заключается в замене уравнений (3.1) на:

$$dx_{i,j} = \eta dW_{i,j} + \begin{cases} F_X \left(\frac{V_{te}^j - V_{int}^j}{R_{i,j}}, V_{te}^j - V_{int}^j, x_{i,j} \right) dt, & V_g^i(t) > 0, \\ 0, & V_g^i(t) = 0, \end{cases}$$

$$R_{i,j} = F_R(V_{te}^j - V_{int}^j, x_{i,j}), i = \overline{1, n}, j = \overline{1, m}$$

где n – количество входов, m – количество нейронов, V_g^i – текущее значение напряжения на i -м входе нейронной сети; V_{te}^j – текущее значение напряжения в обратной связи j -го нейрона; V_{int}^j – напряжение на конденсаторе j -го нейрона; $R_{i,j}$ – значение сопротивления мемристора i -го синапса j -го нейрона; $x_{i,j}$ – состояние мемристора i -го синапса j -го нейрона; η – коэффициент, характеризующий интенсивность шума; $W_{i,j}$ – винеровский процесс, который соответствует i -му мемристору j -го нейрона.

В результате использования стохастических уравнений состояния мемристоров, вся модель нейроморфной сети становится также стохастической.

Рассмотрим сеть, состоящую из одного нейрона ($m = 1$, поэтому индекс j в дальнейшем опускается) с 64 синапсами. Значения параметров: $n = 64$,

$R_{int} = 1 \text{ кОм}$, $C_{int} = 45 \text{ мкФ}$, $V_{te}^+ = 1.5 \text{ В}$, $V_{te}^- = -1.6 \text{ В}$, $V_{te}^0 = 10 \text{ мВ}$, $V_{out}^+ = 2 \text{ В}$, $V_{th} = 3 \text{ мВ}$, $\tau_r = 20 \text{ мс}$, $\tau_s = 2 \text{ мс}$, $\tau_{out} = 10 \text{ мс}$. Данные значения приведены в предыдущем разделе для детерминированной модели нейроморфной сети. Параметры V_{te}^+ , V_{te}^- , V_{te}^0 , τ_r , τ_s определяют амплитуды, продолжительность и задержку между импульсами в обратной связи благодаря которым происходит обучение согласно методу STDP. В качестве модели мемристора используется модель (2.5) (оксид гафния).

В процессе моделирования, в каждой эпохе (равной $\tau_r / 2$ сек) вектор входных сигналов $V_g(t)$ соответствует распознаваемому шаблону (рис. 3.17) или задается случайно (элементы вектора имеют дискретное распределение: $V_g^i = 0 \text{ В}$ с вероятностью 0.81 и $V_g^i = 2 \text{ В}$ с вероятностью 0.19). Для наглядности вектор V_g , который соответствует шаблону, здесь записывается в виде матрицы:

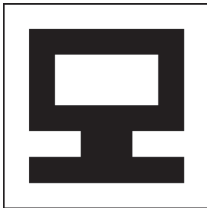


Рис. 3.17. Распознаваемый шаблон

$$V_g = \begin{pmatrix} 0, 0, 0, 0, 0, 0, 0, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 2, 0, 0, 0, 0, 2, 0, \\ 0, 2, 0, 0, 0, 0, 2, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 0, 0, 2, 2, 0, 0, 0, \\ 0, 2, 2, 2, 2, 2, 2, 0, \\ 0, 0, 0, 0, 0, 0, 0, 0 \end{pmatrix} \text{ В}$$

Далее приведены графики поведения характерных напряжений нейронной сети в зависимости от времени (рис. 3.18 и 3.17). Верхние графики соответствуют нулевому значению коэффициента, характеризующего интенсивность шума, то есть фактически детерминированной модели мемристора, а нижние графики – двум разным реализациям при значении $\eta = 0.05$. Пунктирной линией показано, какой сигнал в каждую конкретную эпоху подавался на вход нейронной сети. Значению -1 соответствует шум, а значению 0 – распознаваемый шаблон. Пунктирная линия с точкой отражает значение напряжения на конденсаторе

нейрона, сплошная серая тонкая линия – значение напряжения в обратной связи. Сплошная толстая линия соответствует напряжению на выходе сети. Как только напряжение на конденсаторе достигает определенного порога, происходит активация нейрона: появляется импульс на выходе и серия импульсов в обратной связи.

Во всех случаях начальные условия и процесс обучения были зафиксированы (на вход в сеть подавалась одна и та же последовательность значений). По графикам наблюдается, что сеть успешно обучилась, при этом, за счет наличия стохастического члена в модели мемристора, активации нейрона происходят с некоторыми случайными смещениями во времени. На рис. 3.20 показан процесс изменения переменных состояний мемристоров для пяти различных реализаций. Здесь наблюдается, что стохастическая добавка в модель мемристора приводит к различной скорости адаптации весов: в частности, в первой реализации распознаваемый шаблон начинает просматриваться уже с 1500-й эпохи, а в остальных реализациях – только с 2000-ой эпохи.

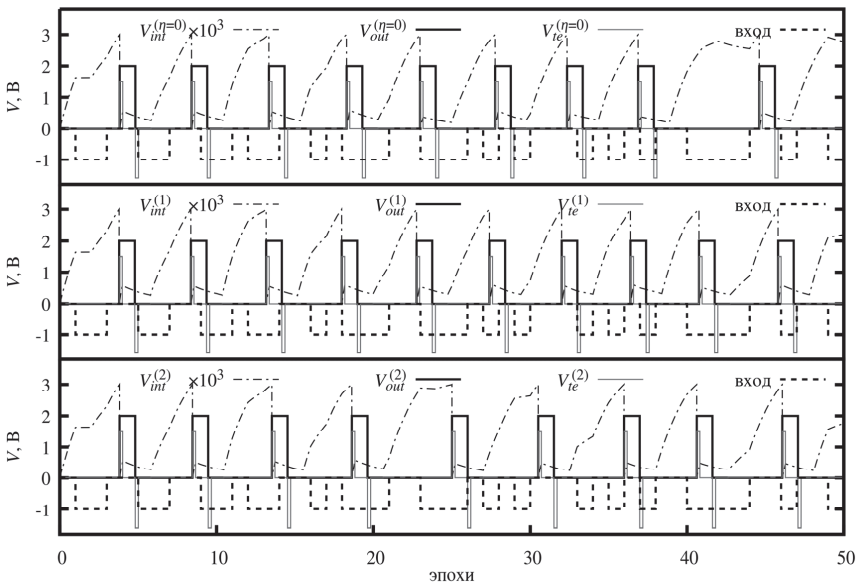


Рис. 3.18. Сравнение поведения параметров нейронной сети в начале обучения при $\eta = 0$ (верхний график) и при $\eta = 0.05$ (нижние графики)

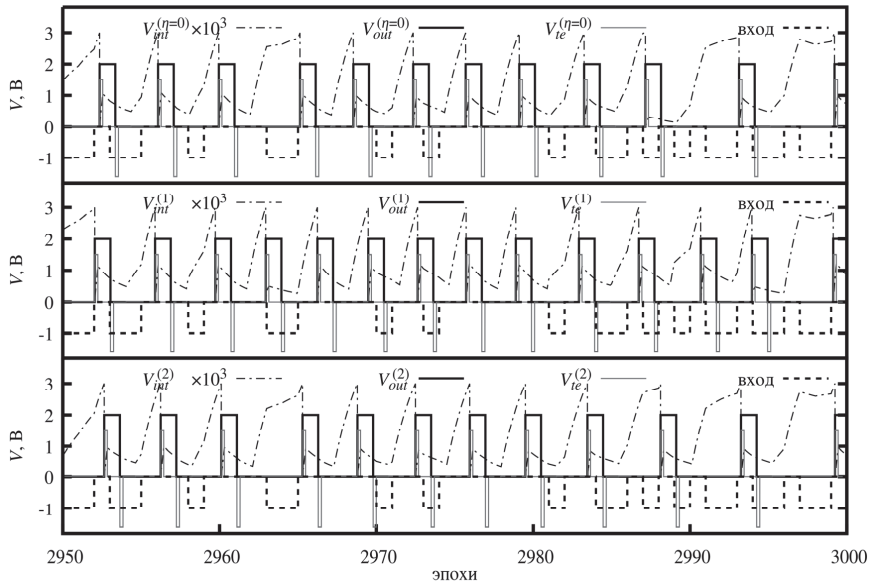


Рис. 3.19. Сравнение поведения основных параметров нейронной сети в конце обучения при $\eta = 0$ (верхний график) и при $\eta = 0.05$ (нижние графики)

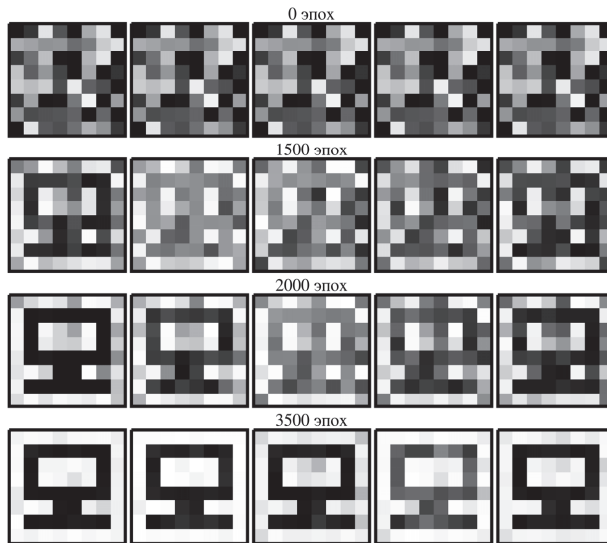


Рис. 3.20. Сравнение изменений весов в разных реализациях

Далее рассматривается задача распознавания двух образцов (3.6) (рис. 3.21). Параметры модели нейроморфной сети остаются прежними, за исключением $V_{th} = 4$ мВ. Так как здесь происходит взаимодействие между

нейронами, дополнительно задается коэффициент подавления $\alpha = 0.1$ (в момент активации одного нейрона потенциал другого нейрона сбрасывается в соответствии с коэффициентом подавления).

Как и ранее, в каждой эпохе вектор входных сигналов $V_g(t)$ соответствует распознаваемым шаблонам (рис. 3.21) или задается случайно (элементы вектора имеют дискретное распределение: $V_g^i = 0\text{В}$ с вероятностью 0.8 и $V_g^i = 2\text{В}$ с вероятностью 0.2).



Рис. 3.21. Распознаваемые шаблоны

На рис. 3.22 и 3.23 приведены графики зависимости напряжений от времени в начале и в конце процесса обучения для двух различных реализаций. На рис. 3.24 продемонстрировано изменение переменных состояний мемристоров для пяти реализаций.

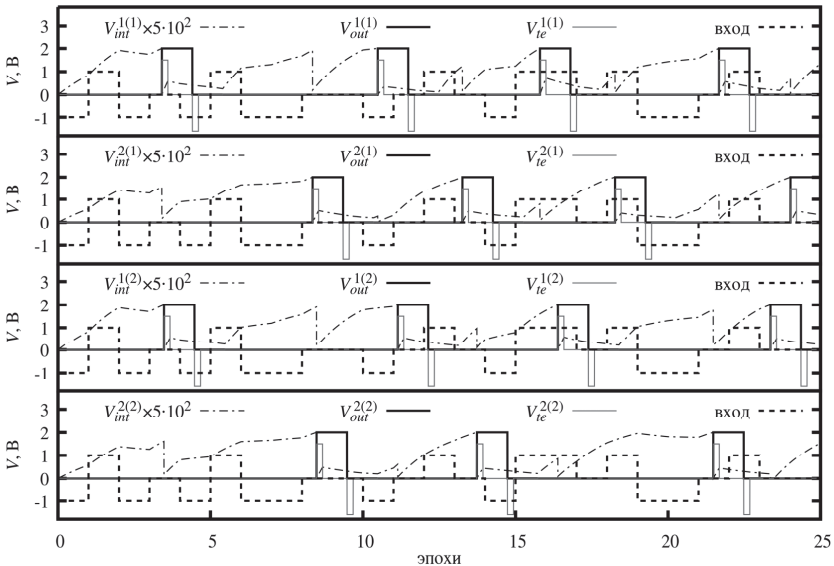


Рис. 3.22. Сравнение поведения основных параметров нейронной сети в начале обучения в двух разных реализациях

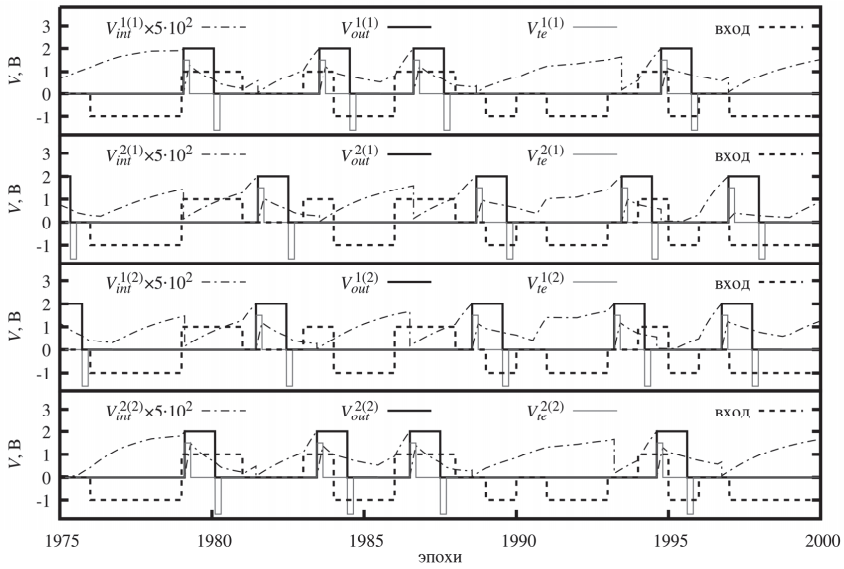


Рис. 3.23. Сравнение поведения основных параметров нейронной сети в конце обучения в двух разных реализациях

Пунктирной линией показано, что в каждую конкретную эпоху подавалось на вход нейронной сети. Значению -1 соответствует шум, значению 0 – первый распознаваемый шаблон, значению 1 – второй распознаваемый шаблон.

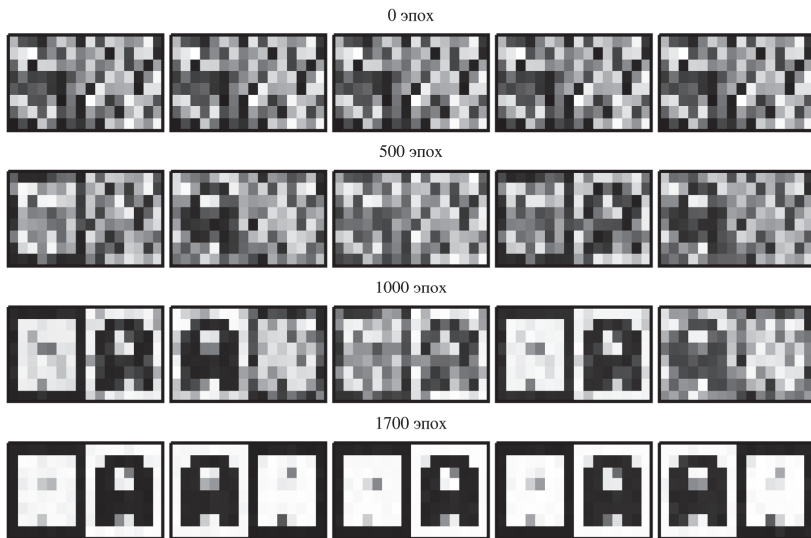


Рис. 3.24. Сравнение изменений весов двух нейронов в разных реализациях

В результате имитационного моделирования работы нейроморфной сети получено, что сеть успешно обучилась распознавать заданные шаблоны при значениях параметров, подобранных в рамках детерминированной модели. Это говорит о том, что настройку сети возможно выполнять без учета стохастической природы мемристивных элементов. В первую очередь, аддитивный шум сказывается на то, каким образом шаблоны распределяются по нейронам и в меньшей степени на задержки в выходных импульсах.

Главной характеристикой любой нейронной сети является точность. Обычно для её оценки в сеть подается на вход тестовый набор данных и определяется процент правильных ответов. Отличительной особенностью рассматриваемой сети является наличие внутреннего состояния, которое изменяется в процессе функционирования, поэтому классический подход тестирования не совсем корректен. В связи с этим для оценивания точности выполняется усреднение по некоторому временному окну в процессе обучения сети: подсчитывается сколько раз сеть правильно отреагировала на входные данные. Используемый размер окна – 100 эпох. Так же вопрос вызывает то, что шаблоны в процессе обучения случайным образом могут перераспределяться по нейронам, поэтому заранее перед процессом обучения определить какой из нейронов будет отвечать за какой шаблон нельзя. Для решения этого вопроса на каждой эпохе проверяется, к какому нейрону больше всего подходит шаблон: максимизируется скалярное произведение между входным вектором и текущим вектором значений синаптических весов каждого нейрона.

На рис. 3.25 и 3.26 показаны зависимости точности от номера эпохи для двух рассмотренных ранее сетей. Кривые соответствующие нулевому значению коэффициента шума фактически получены с использованием детерминированной модели. Для ненулевых значений коэффициента шума на графиках приведены по пять разных кривых, соответствующих разным реализациям. Видно, что при небольшом значении η сеть в ряде реализаций обучается быстрее, чем при отсутствии шума. Однако с увеличением η

процесс обучения теряет стабильность и перестает сходиться, о чем свидетельствует сильный разброс в графиках при $\eta = 0.25$.

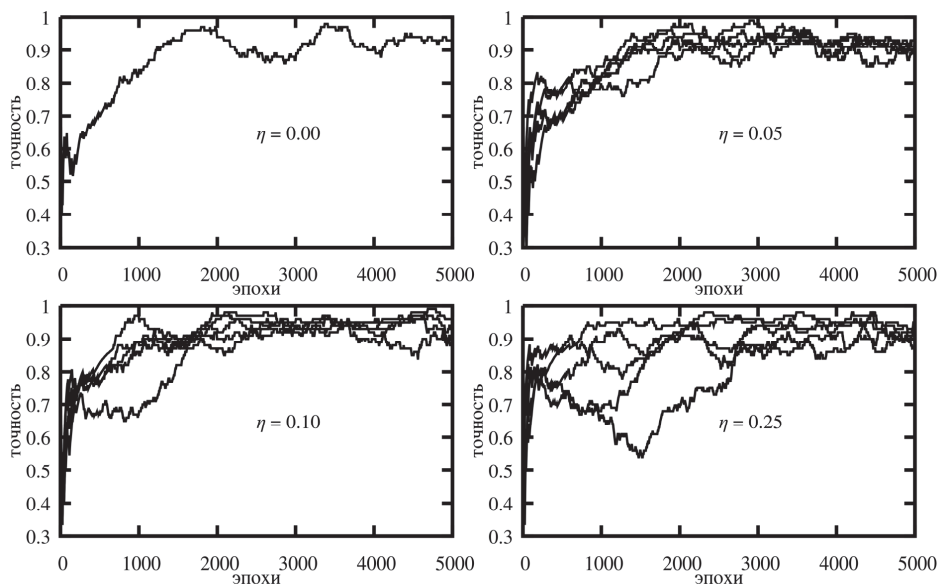


Рис. 3.25. Зависимость точности распознавания от номера эпохи в случае с одним образцом

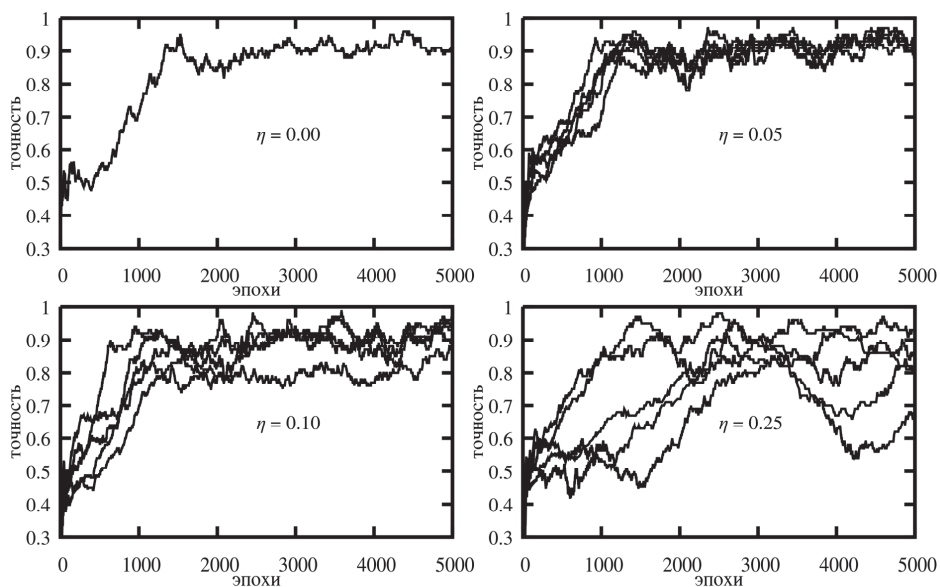


Рис. 3.26. Зависимость точности распознавания от номера эпохи в случае с двумя образцами

Далее приведем параметры сети предназначенной для распознавания пяти изображений (рис. 3.14) с использованием модели мемристора (2.4) (оксид титана): $\eta = 0.05$, $R_{int} = 200$ Ом, $C_{int} = 45$ мкФ, $V_{te}^+ = 0.7$ В, $V_{te}^- = -0.9$ В, $V_{te}^0 = 10$ мВ, $V_{out}^+ = 2$ В, $V_{th} = 9$ мВ, $\tau_r = 3$ мс, $\tau_s = 50$ мкс, $\tau_{out} = 1.5$ мс, (распределение шума, который подается на вход в сеть в процессе обучения: $V_g^i = 0$ В с вероятностью 0.73 и $V_g^i = 2$ В с вероятностью 0.27).

На рис. 3.27 показан процесс адаптации синаптических весов к распознаваемым образцам. Цвет соответствует значению переменной состояния соответствующего мемристора: чем темнее, тем проводимость больше; чем светлее, тем меньше. В начальный момент времени все веса инициализируются случайными значениями, и в процессе работы сети постепенно изменяются. Примерно с 1200-ой эпохи начинают просматриваться шаблоны, распознаванию которых обучается сеть: происходит запоминание информации нейронной сетью. Длительность одной эпохи – 1.5 мс.

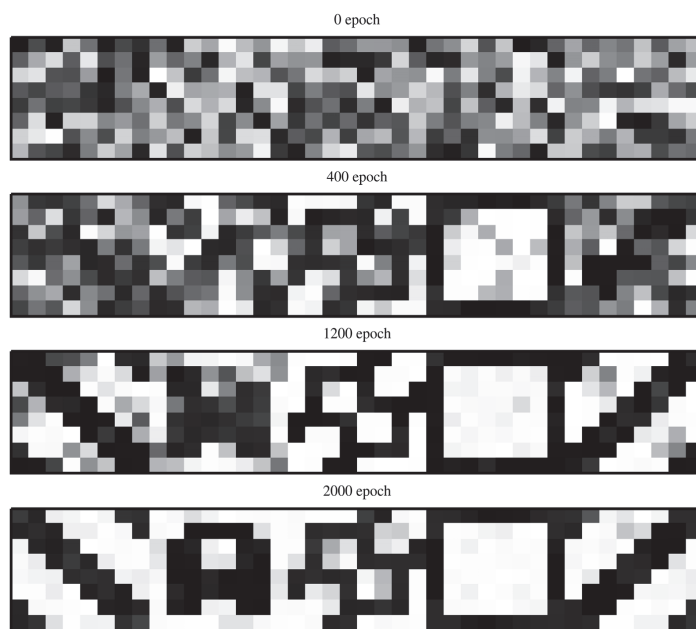


Рис. 3.27. Изменение синаптических весов в процессе обучения нейроморфной сети

На примере распознавания образов с использованием нейроморфных сетей с одним и двумя нейронами показано, что скорость адаптации весов сети к распознаваемым шаблонам, а также распределение шаблонов по нейронам, зависят от стохастических особенностей мемристивных элементов. При этом настройка параметров сети может проводиться без учета стохастической динамики переключения мемристоров. При относительно невысоком уровне шума нейронные сети с параметрами, подобранными с помощью детерминированной математической модели, успешно обучаются распознавать заданные шаблоны.

3.3. Интервальная модель нейронной сети

В данном разделе предлагается использовать интервальные модели мемристоров при имитационном моделировании работы нейроморфной сети. В процессе моделирования каждую эпоху параметры мемристоров задаются случайным образом из подобранных интервалов. Такой подход позволяет обойтись без применения стохастического математического аппарата. В качестве модели мемристора используется модель (2.6) с параметрами подобранными для LiNbO_3 : $I_{\max} = 100$ мА, $D = 4 \times 10^{-7}$ м, $R_{on} \in [14.46, 29.46]$ Ом, $R_{off} \in [209.6, 210.8]$ Ом, $V_n \in [-5.2, -4.66]$ В, $V_p \in [2.88, 3.04]$ В, $\mu_v \in [3.77 \times 10^{-14}, 5.03 \times 10^{-14}]$, $\delta \in [-0.01, 0.012]$.

Выполняется имитационное моделирование работы сети, предназначенной для распознавания двух шаблонов (3.6) (рис. 3.21). Значения параметров сети: $n = 64$ – количество пикселей в одном шаблоне; $m = 2$ – количество шаблонов; $R_{int} = 14.3$ Ом, $C_{int} = 261.5$ мФ – сопротивление и емкость, определяющие свойства нейронов; $V_{te}^+ = 3.84$ В, $V_{te}^- = -5.72$ В, $V_{te}^0 = 20$ мВ – амплитуды напряжений для изменений синаптических весов (проводимости мемристоров); $V_{th} = 15$ мВ – порог активации нейрона; $\tau_p = 0.83$ с, $\tau_s = 41.5$ мс, $\tau_{out} = 415$ мс – длительности импульсов, участвующих в процессе обучения сети.

На рис. 3.28 показан процесс изменения переменных состояний мемристоров для пяти различных реализаций. В каждой реализации в каждую эпоху параметры мемристора задавались случайным образом, при этом начальное состояние мемристоров, а также последовательность демонстрации сети распознаваемых паттернов оставались неизменными. Здесь наблюдается, что адаптация весов к распознаваемым шаблонам происходит с разной скоростью и шаблоны распределяются по нейронам случайным образом, что соответствует результатам, полученным в предыдущем разделе.

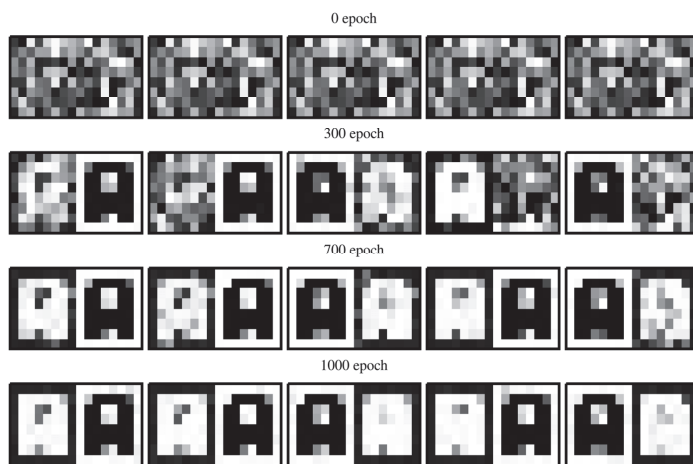


Рис. 3.28. Сравнение изменений весов в разных реализациях

На рис. 3.29 показаны зависимости точности распознавания от номера эпохи для разных реализаций.

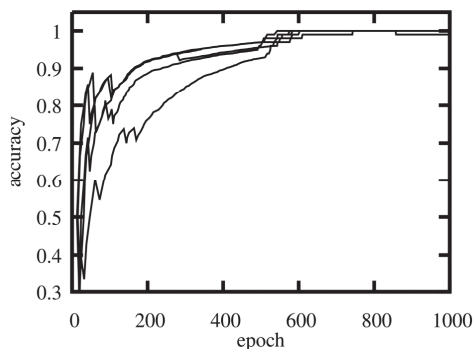


Рис. 3.29. Сравнение точности распознавания в разных реализациях

Представленные графики согласуются с аналогичными графиками на рис. 3.25 и 3.26. Это говорит о том, что предложенный подход пригоден для моделирования нейроморфной сети, при этом здесь нет необходимости решать стохастические ОДУ. Кроме этого, отметим, что используемая интервальная модель мемристора является грубой (полученные интервальные оценки экспериментальных данных во втором разделе являются достаточно широкими), однако, несмотря на это сеть успешно обучилась. Это дополнительно подтверждает универсальность разработанной математической модели сети.

3.4. Параллельная реализация модели нейронной сети

Имитационное моделирование позволяет выполнить настройку сети на уровне математической модели, и впоследствии использовать полученные параметры непосредственно на производстве, что является важным. Модель сети задается в виде системы обыкновенных дифференциальных уравнений, которая может состоять из десятков и сотен тысяч уравнений. Естественным образом возникает потребность в эффективной и параллельной реализации соответствующей имитационной модели. В качестве технологии для распараллеливания вычислений предлагается использовать OpenMP (Open Multi-Processing) [54], так как она позволяет достаточно легко создавать многопоточные приложения на различных языках программирования. OpenMP – это набор директив компилятора, библиотечных процедур и переменных окружения, которые предназначены для программирования многопоточных приложений на многопроцессорных системах с общей памятью [55].

Технология OpenMP начинает свою историю с 1997 г. и за прошедшее время зарекомендовала себя как очень удобная технология для создания легко переносимых многопоточных приложений.

Как правило, для высоконагруженных программ справедливо утверждение, что 90% времени работы приходится на 10% программного кода, и эти 10% являются циклами. В самом простом случае чтобы

распараллелить цикл, итерации которого независимые, достаточно перед циклом написать всего одну строчку: `#pragma omp parallel for`, и после этого итерации будут выполняться на разных вычислительных ядрах. А если, например, в цикле еще происходит суммирование каких-то значений в общую переменную `sum`, то необходимо добавить директиву `reduction(+:sum)`. Это демонстрирует простоту и удобство данной технологии.

В процессе имитационного моделирования выполняется численное решение системы ОДУ, заключающееся в многократном вычислении правой части системы, что является наиболее трудозатратой операцией. Так как в рамках одного вычисления правой части уравнения являются независимыми, то здесь удобно применить технологию OpenMP для распараллеливания циклов соответствующих индексам i и j в описании модели.

Эффективность распараллеливания оценивается на задаче моделирования процесса обучения сети распознаванию набора из пяти изображений [56] (рис. 3.30) размера 128 на 128 пикселей, которая приводит к решению 81930 ($128 \times 128 \times 5 + 2 \times 5$) дифференциальных уравнений. Вычисления производились на процессоре Intel(R) Xeon(R) CPU E5-2620. В качестве модели мемритора использовалась интервальная модель (2.6) с параметрами, соответствующая LiNbO_3 .

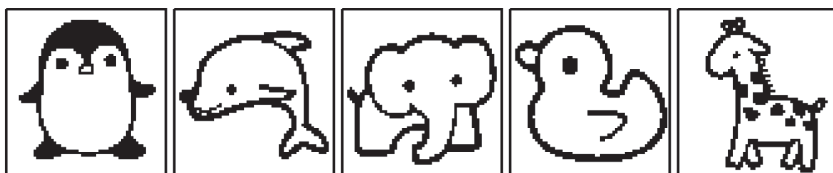


Рис. 3.30. Распознаваемые изображения

В процессе имитационного моделирования подобны следующие значения параметров модели: $R_{int} = 10.9$ Ом, $C_{int} = 4.1$ мФ, $V_{ie}^+ = 1.55$ В, $V_{ie}^- = -1.6$ В, $V_{ie}^0 = 10$ мВ, $V_{out}^+ = 2$ В, $V_{th} = 2.5$ мВ, $\tau_r = 15$ мс, $\tau_s = 1$ мс, $\tau_{out} = 1$ мс, $\alpha = 0.1$.

На рис. 3.31 показан процесс адаптации синаптических весов к распознаваемым образцам.

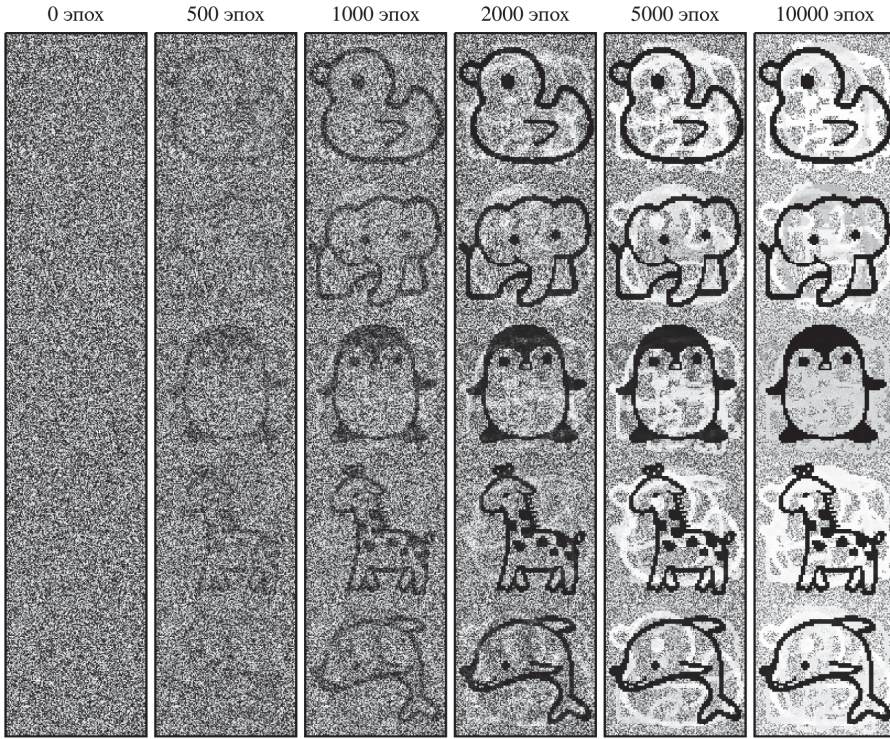


Рис. 3.31. Изменение состояний мемристоров в процессе обучения

Цвет соответствует значению переменной состояния соответствующего мемристора: чем темнее, тем проводимость больше; чем светлее, тем меньше. Соответствие переменных состояний распознаваемым образцам говорит о том, что сеть успешно обучилась и запомнила заданные изображения.

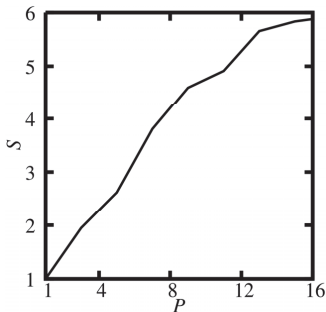


Рис. 3.32
Зависимость ускорения
от количества вычислительных
потоков

На рис. 3.32 показана зависимость ускорения от количества вычислительных потоков ($S = T_1/T_p$, где T_1 – время работы последовательной версии реализации; T_p – время работы параллельной версии реализации с использованием P вычислительных потоков). За счет распараллеливания удалось получить шестикратное ускорение вычислений.

ГЛАВА 4

АНАЛОГОВАЯ РЕАЛИЗАЦИЯ КЛАССИЧЕСКИХ НЕЙРОННЫХ СЕТЕЙ

Присущее мемристорным кроссбарам свойство естественной параллелизации матрично-векторных операций создает возможности для их эффективного использования в аналоговой реализации классических нейронных сетей. Аналоговые вычисления производятся на порядки быстрее по сравнению с вычислениями на центральном процессоре и на графических ускорителях. Кроме того, существенно ниже энергозатраты на проведение математических операций. При этом существенной особенностью аналоговых вычислений является небольшая точность. Корпорация Hewlett-Packard уже создала в «железе» мемристорный кроссбар, на котором реализована операция матрично-векторного умножения [36, 37]. На данный момент точность задания матрицы ограничивается 6 битами (для каждого мемристора выделяется 64 уровня проводимости). В связи с этим актуальным является исследование зависимости качества работы классических нейронных сетей от точности задания весов.

Рассматриваемый подход предполагает разделение этапов обучения и анализа данных. Как правило, на этапе обучения используются градиентные методы оптимизации, реализацию которых целесообразно выполнять на центральном процессоре (или на графических ускорителях). Этап анализа данных состоит в прямом распространении сигнала по сети, что, учитывая возможности параллельного выполнения матрично-векторных умножений, идеально ложится на архитектуру мемристорного кроссбара. Таким образом, этап анализа данных эффективно реализуется в аналоговом режиме.

При таком подходе возникают две основные проблемы. Первая связана с переходом от непрерывного диапазона изменения синаптических весов к дискретному набору значений, обеспечиваемых мемристором. В наиболее радикальном варианте, когда мемристор может находиться в двух

состояниях, высокоомном и низкоомном, речь идет о бинаризации синаптических весов. Ясно, что такая дискретизация может вносить погрешности в результаты анализа данных. Поэтому актуальной задачей является выявление зависимости погрешности дискретизации от количества дискретных уровней для различных наборов данных.

Вычисления в каждом слое сети включают перемножение матрицы весов на входной вектор, прибавление вектора смещения и применение нелинейной функции активации. Все эти действия возможно выполнить в аналоговом представлении.

Нейронная сеть заранее обучается с использованием центрального процессора или графических ускорителей и далее её веса размещаются в соответствующих мемристорных кроссбарах.

Рассматриваются две сверточные нейронные сети. Первая состоит из двух сверточных слоев (conv), одного слоя подвыборки (maxpool) и двух полносвязанных слоев (dense) (рис. 4.1), а вторая из четырех сверточных слоев, двух слоев подвыборки и двух полносвязанных слоев (рис. 4.2).

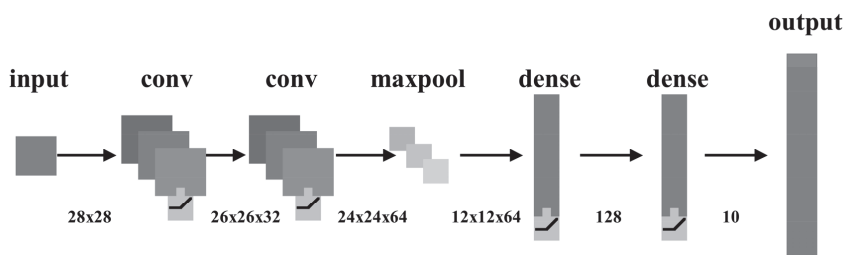


Рис. 4.1. Архитектура сверточной нейронной сети (MNIST)

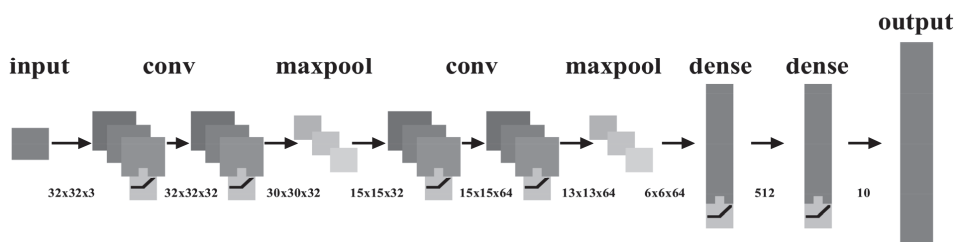


Рис. 4.2. Архитектура сверточной нейронной сети (CIFAR_10)

Свёрточный слой (conv) состоит из набора ядер (рис. 4.3). Каждое ядро это матрица (в общем случае тензор) небольшого размера (в представленных сетях, размером 3×3 и $3 \times 3 \times 3$) которая «накладывается» с перехлестом на входные данные с последующим перемножением соответствующих элементов и суммированием. Например, такие операции как размытие изображения или выделение контуров являются сверткой. Если представить входные данные в виде вектора, а ядра свертки в виде больших разреженных матриц, то получение результата сверточного слоя сведется к перемножению матрицы на вектор, что в свою очередь эффективно реализуется на мемристорных кроссбарах.

Слой подвыборки (maxpool) применяется для уменьшения размера данных (рис. 4.3). В рассмотренных сетях из каждых четырех промежуточных значений оставляется только одно – максимальное. Здесь подразумевается, что операция выбора максимального элемента из нескольких может быть реализована на аналоговом уровне. Отметим, что если для уменьшения размера данных использовать не выбор максимального элемента, а усреднение, тогда данный слой можно представить как сверточный слой, в котором ядра покрывают входные данные без перехлеста.

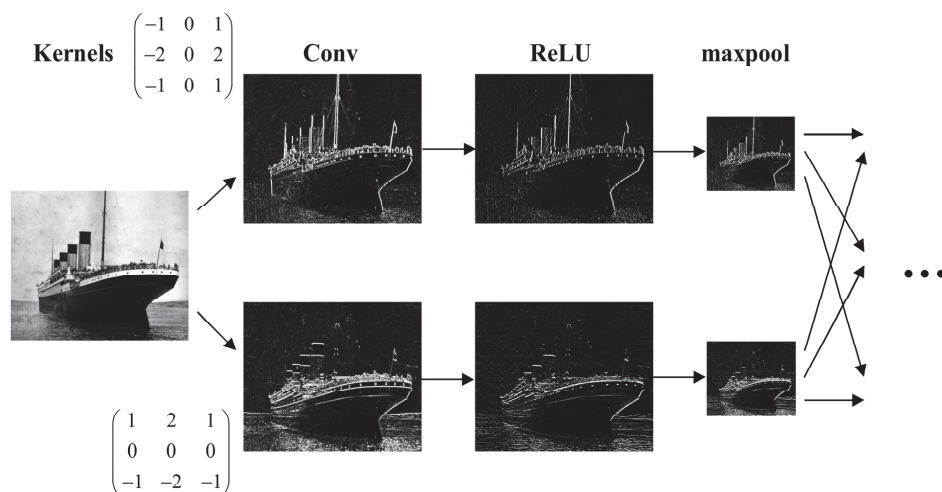


Рис. 4.3. Часть сверточной нейронной сети

К результату каждого слоя (кроме слоя подвыборки) поэлементно применяется активационная функция ReLU (рис. 4.4):

$$f(x) = \max(0, x)$$

Данная функция представляется на железном уровне в виде полупроводникового диода.

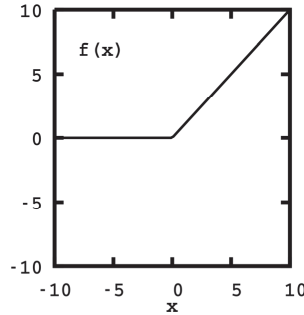


Рис. 4.4. Активационная функция ReLU

В целом все элементы рассмотренных нейронных сетей возможно реализовать на аналоговых электронных компонентах.

Для обучения и работы с сетями использовалась библиотека Keras, написанная на языке Python, которая является надстройкой над фреймворками DeepLearning4j, TensorFlow и Theano [57].

Первая нейронная сеть (рис. 4.1) была обучена на наборе данных MNIST состоящем из изображений рукописных цифр (рис. 4.5) [58]. Вторая нейронная сеть (рис. 4.2) обучалась на CIFAR_10, состоящем из самолетов, автомобилей, животных и тд. (рис. 4.6) [59]. Точность распознавания на тестовых выборках обученной нейронной сети составила: для MNIST – 99.16% и для CIFAR_10 – 74.86%. Отметим, что для CIFAR_10 возможна реализация с более высокой точностью, однако целью данной работы является не достижение точности искусственной нейронной сети, а оценка погрешности вследствие дискретизации весов.



Рис. 4.5. Набор данных MNIST

Далее была выполнена дискретизация весов: начиная с двух значений (1 бит точности) и до 1024 значений (10 бит точности). Дискретизация выполнялась в рамках каждого слоя отдельно. На рис. 4.7 и 4.8 показаны зависимости качества распознавания от количества бит точности задания весов сети. Пунктирной линией показана исходная точность распознавания.



Рис. 4.6. Набор данных CIFAR_10

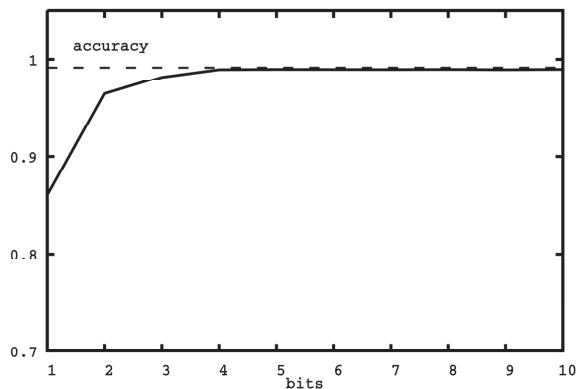


Рис. 4.7. Зависимость качества нейронной сети от точности задания весов (MNIST)

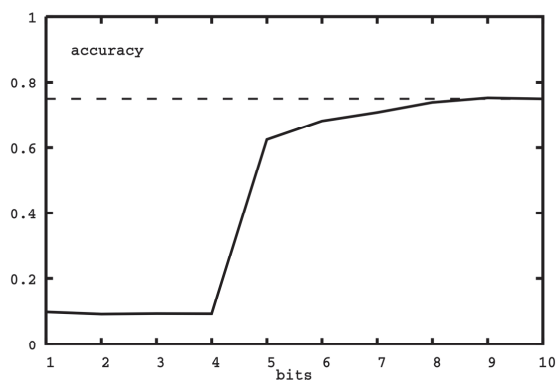


Рис. 4.8. Зависимость качества нейронной сети от точности задания весов (CIFAR_10)

Для сети, обученной на MNIST'е, для достижения приемлемой точности (98–99%) требует 3–4 бита точности. А для нейронной сети, обученной на CIFAR_10, нужно 6–8 бит. Таким образом, можно полагать, что 6-ти битные мемристоры созданные корпорацией Hewlett-Packard, возможно использовать для аналоговой реализации рассмотренных нейронных сетей.

ГЛАВА 5

ПРИМЕНЕНИЕ ИНТЕГРАЦИОННОЙ ПЛАТФОРМЫ ДЛЯ МНОГОМАСШТАБНОГО МОДЕЛИРОВАНИЯ НЕЙРОМОРФНЫХ СИСТЕМ

5.1. Принципы формирования информационной поддержки многомасштабного моделирования физических явлений и процессов

В данной работе при построении многомасштабной модели для имитации работы нейроморфных систем используются подходы, разработанные ранее и представленные в [60, 61]. В их основе лежит идея о том, что спецификация информационных объектов предметной области, может быть сформулирована в терминах конечно-множественных представлений [62] которые, в частности, могут быть реализованы посредством доменных моделей [63, 64] – конечных наборов (множеств) данных с поддержкой операций по их обработке. В статье [61] даны понятия и определения, применяемые при описании информационной технологии многомасштабного моделирования. Вводится понятие «базовая модель-композиция (БК)» для ее описания применяется теоретико-множественный аппарат. Используется обозначение базовой модели – композиции: MC_i^j , здесь i – номер масштабного уровня, а j – номер данной БК на этом масштабном уровне [60, 61].

Как было показано в работах [60, 61] БК являются информационными аналогами базовых физико-математических моделей, которые применяются в вычислительном процессе для решения конкретных задач на своем масштабном уровне. Их можно представить в виде таблиц данных разного структурного типа (входных/выходных данных, моделей и алгоритмов, другое). Экземпляры БК (таблицы с конкретными данными) представляются

объектами классов –наследников БК в объектно-ориентированном языке программирования. Они хранятся в виде документов в документно-ориентированной базе данных [65]. Многомасштабная композиция представляет собой коллекцию, состоящую из сгруппированных документов с иерархической структурой, отражающей последовательность передачи данных в общем вычислительном процессе.

Покажем, как при решении задачи о применении методов многомасштабного моделирования для имитации работы нейроморфной сети, с помощью данной технологии из конкретных базовых моделей-композиций составляются многомасштабные композиции (МК) – информационные аналоги многомасштабных моделей, передающие содержание многомасштабных вычислительных процессов.

5.2. Интеграционная платформа для моделирования работы нейроморфной сети

Интеграционная платформа для многомасштабного моделирования нейроморфной сети, объединяет информационные потоки на различных масштабных уровнях – на квантово-механическом и наноуровнях, на уровне элементов резистивной памяти, на уровне нейроморфной сети, на уровне имитации обучения нейроморфной сети по прецедентам. Для более детального описания моделируемого вычислительного уровня введем дополнительные уровни.

Общее представление о распределении базовых моделей-композиций по масштабным уровням, задействованных в вычислительном процессе показано в Таблице 5.1.

Таблица 5.1

№ маш. уровня	Обозначение и название базовой модели-композиции	Название масштабного уровня
0	MC_0^i «АТОМ A_0^i »	Уровень химич.элементов
1	MC_1^1 «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» MC_1^2 «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА»	Квантово-механический уровень
2	MC_2^1 «АТОМНЫЙ КЛАСТЕР-СТАТИКА» MC_2^2 «АТОМНЫЙ КЛАСТЕР-ДИНАМИКА»	наноуровень
3	MC_3^1 «Модель генерации/рекомбинации ионов кислорода и кислородных вакансий (GR-model)» MC_3^2 «Модель дрейфа-диффузии ионов кислорода (DD-model)» MC_3^3 «Модель силового поля (E-model)» MC_3^4 «Модель переноса электронов (электрического тока) (J-model)» MC_3^5 «Модель теплопереноса (HT-model)»	Уровень элемента резистивной памяти (мемристора)
4	MC_4^1 «Фитинг моделей.ВАХ» MC_4^2 «Схемотехническое представление»	Уровень формирования нейроморфной сети
5	MC_5^1 «Нейросетевое моделирование». Обучение по прецедентам	Логический уровень

Покажем, как организуется вычислительным процесс, имитирующий работу нейроморфной сети, в основе которой лежит мемристивный элемент на базе оксида гафния.

На нулевом масштабном уровне с помощью базовой модели-композиции MC_0^i «АТОМ A_0^i » (i – номер элемента в Таблице Менделеева (ТМ)) задаются основные данные по конкретным химическим элементам, входящим в состав соединений, участвующих в вычислительном процессе

(номер химического элемента из ТМ, масса атома, заряд ядра, радиус атома, электронная конфигурация, структура решетки, другое). В нашем случае формируются экземпляры базовой модели-композиции MC_0^{72} «АТОМ H_2^{72} » и MC_0^{16} «АТОМ O_0^{16} ».

К первому масштабному уровню отнесены базовые модели-композиции MC_1^4 «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» и MC_1^2 «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». С их помощью данные, полученные с нулевого масштабного уровня из MC_0^{72} «АТОМ H_2^{72} » и MC_0^{16} «АТОМ O_0^{16} », передаются на квантово-механический уровень первоначально в MC_1^4 «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА», где используя знания о химическом составе и кристаллографической структуре, с помощью простых физико-математических моделей (ионно-атомных радиусов, другое) определяется кристаллохимическая структура соединения (метрические параметры кристаллической решетки, координаты базисных атомов, другое). Данная БК программно реализована в двух расчетных модулях (модель «Плотная упаковка» и программный комплекс Materials Studio [66].

Далее, полученные в ходе вычислительного процесса данные передаются в базовую модель-композицию MC_1^2 «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». Данная БК программно реализована в двух расчетных модулях (программный комплекс VASP [67] и пакет программ с открытым кодом Quantum Espresso [68]. Здесь на базе теории функционала электронной плотности решается задача по определению итоговых значений параметров кристаллической решетки, рассчитывается электронная плотность [69], полная энергия для заданной конфигурации базисных атомов. Далее проводятся расчеты констант упругости, поляризация, фононное рассеяние, рассчитываются энергетические барьеры, энергия активации, другое.

На втором масштабном уровне, используя данные, полученные на предыдущем масштабе, производится выбор потенциалов межатомного

взаимодействия, с учетом тип связи моделируемого соединения (в данном случае оксида гафния). Например, используется модель погруженного атома (embedded atom model, EAM [70], либо потенциал межатомного взаимодействия MEAM [71]. При помощи базовой модели-композиции **MC₂¹** «АТОМНЫЙ КЛАСТЕР-СТАТИКА» проводится идентификация параметров потенциала межатомного взаимодействия. Далее потенциал с идентифицированными параметрами используется в ходе Молекулярно-динамического моделирования в рамках базовой модели-композиции **MC₂²** «АТОМНЫЙ КЛАСТЕР-ДИНАМИКА». Ранее полученные данные о координатах позиций атомов кислорода, рассчитанные данные о энергии активации, энергии рекомбинации, энергетические барьеры миграции и другое поступают на третий масштабный уровень. Туда же поступают дополнительные данные, такие как значения потенциалов на электродах, коэффициенты, учитывающие вклад поляризации света в энергетические барьеры и другое.

На третьем масштабном уровне сформирована математическая модель образования/разрушения проводящих каналов (филаментов) в мемристорных элементах на основе оксидных пленок. Ей соответствует «Композитный модуль расчета мемристора». Необходимо отметить, что в мире весьма активно ведутся работы по исследованию возможностей проектирования и оптимизации ячеек памяти на основе оксидов металлов. Так в статье [72] представлен электротермический симулятор, позволяющий исследовать физику и потенциал ячеек резистивной памяти на основе SiOx с произвольным доступом (RRAM). В нашей работе, в контексте многомасштабного подхода, используется модельное представление, включающее в себя пять взаимосвязанных подмоделей, каждой из которых поставлены в соответствие базовые модели-композиции: **MC₃¹** «Модель генерации/рекомбинации ионов кислорода и кислородных вакансий (GR-model)», **MC₃²** «Модель дрейфа-диффузии ионов кислорода (DD-model)»,

MC_3^2 «Модель силового поля (E-model)», MC_3^4 «Модель переноса электронов (электрического тока) (J-model)», MC_3^5 «Модель теплопереноса (HT-model)». Из пяти БК формируется Композиция для расчета свойств мемристора $K_3^{3,1;3,2;3,3;3,4;3,5}$. Здесь нижний индекс $i=3$, так как все БК, образующие Композицию с третьего уровня, верхние индексы ($3,1; 3,2; 3,3; 3,4; 3,5$) через точку с запятой обозначают какие именно БК используются.

На 4-м масштабном уровне многомасштабной модели для имитации работы нейроморфной сети представлена базовая модель-композиция MC_4^1 «Фитинг моделей. ВАХ» и БК MC_4^2 «Схемотехническая». На вход MC_4^1 подаются сведения (графики) по вольтамперным характеристикам (ВАХ) моделируемого соединения с нижнего масштабного уровня, либо сведения о модели и начальном наборе ее параметров. На выходе получаем уточненные значения параметров модели. Далее, полученные данные передаются в БК MC_4^2 «Схемотехническая», туда же поступают данные по параметрам электронной схемы.

На 5-м масштабном уровне проводится моделирование работы аналоговой нейроморфной сети MC_5^1 «Нейросетевое моделирование». Данные, рассчитанные на нижних масштабных уровнях (электронная схема, уточненные ВАХ, уточненные названия параметров модели, другое), а также параметры нейроморфной сети и наборы паттернов распознавания подаются на вход MC_5^1 «Нейросетевое моделирование». С помощью данной модели проводится имитация обучения нейроморфной сети. Более подробно работа модели данного масштабного уровня описана в главе 4.

Для решения практических задач многомасштабного моделирования в области микроэлектроники актуальным является разработка программных комплексов, развернутых на высокопроизводительных кластерах, которые предоставляют исследователям возможность проводить многоуровневые и

мультифизические расчеты, используя как собственные программные разработки, так и пакеты прикладных программ.

В данной работе представлен программный комплекс для информационной поддержки мультифизических и многоуровневых расчетов, базирующийся на технологии многомасштабного моделирования, описанной в [60]. Согласно данной технологии физико-математическим моделям, отнесенным к соответствующим масштабным уровням, поставлены в соответствие информационные структуры – базовые модели-композиции (композиционные объекты). Для описания базовых моделей-композиций и технологии построения многомасштабных композиций используется теоретико-множественный аппарат [60, 61], позволяющий передать вычислительную сущность соответствующих математических. Базовые модели-композиции, классифицированные с учетом масштабной иерархии, применяются для построения композиций и многомасштабных композиций – вычислительных аналогов многомасштабных моделей.

Необходимо отметить, что при построении программного комплекса для решения конкретных задач, используются как программные модули собственной разработки, в которых реализованы оригинальные численные решения созданных физико-математических моделей, так и готовые программные модули. В его основе лежит интеграционная платформа для многомасштабного моделирования, которая объединяет информационные потоки на разных масштабных уровнях. При решении конкретной задачи, такой как имитационное моделирование работы нейроморфной сети или оптимизация датчиков магнитных и температурных полей на основе наноразмерных гетероструктур «кремний на изоляторе» [73], выделяются конкретные уровни моделирования, которые необходимо задействовать, строится многомасштабная композиция – информационный аналог многомасштабной физико-математической модели. Для программной реализации на базе интеграционной платформы из имеющихся программных модулей формируется вычислительный комплекс.

Для реализации сценария предлагается схема, описанная на рисунке ниже. Схема представляет собой исполняемый BPMN [74] процесс, организующий последовательность выполнения расчетов интегрированными модулями, задание входных данных и обмен данными между модулями в процессе выполнения.

Рассмотрим этапы расчета из (рис. 5.1) и их отображение на сценарии:

1. Этап задания входных данных включает ввод необходимой информации о кристаллической структуре вещества. На практике это подразумевает загрузку файла в формате CIF [75]. Данный CIF файл можно получить из открытых баз данных [76], или экспортировать из используемой программы, используемой для моделирования материала на атомном уровне. Формат CIF (Crystallography Information File) [75] представляет собой один из наиболее широко используемых универсальных форматов передачи кристаллографической информации. В определенной степени он стандартизован, однако многие средства используют собственные или нестандартные расширения данного формата. Поэтому при импорте из формата берется лишь необходимый минимум информации для построения базисной ячейки кристалла и основные параметры. Таковыми являются:

- форма кристаллической решетки (геометрия и пространственная группа);
- размеры элементарной ячейки и углы между базисными векторами;
- координаты базисных атомов;
- набор операций симметрии, соответствующий пространственной группе.

Далее, по координатам базисных атомов и набору операций симметрии в реализованной утилитной библиотеке к сервису сценариев генерируется полный набор атомов элементарной ячейки, которая в дальнейшем используется в квантово-механических расчетах.

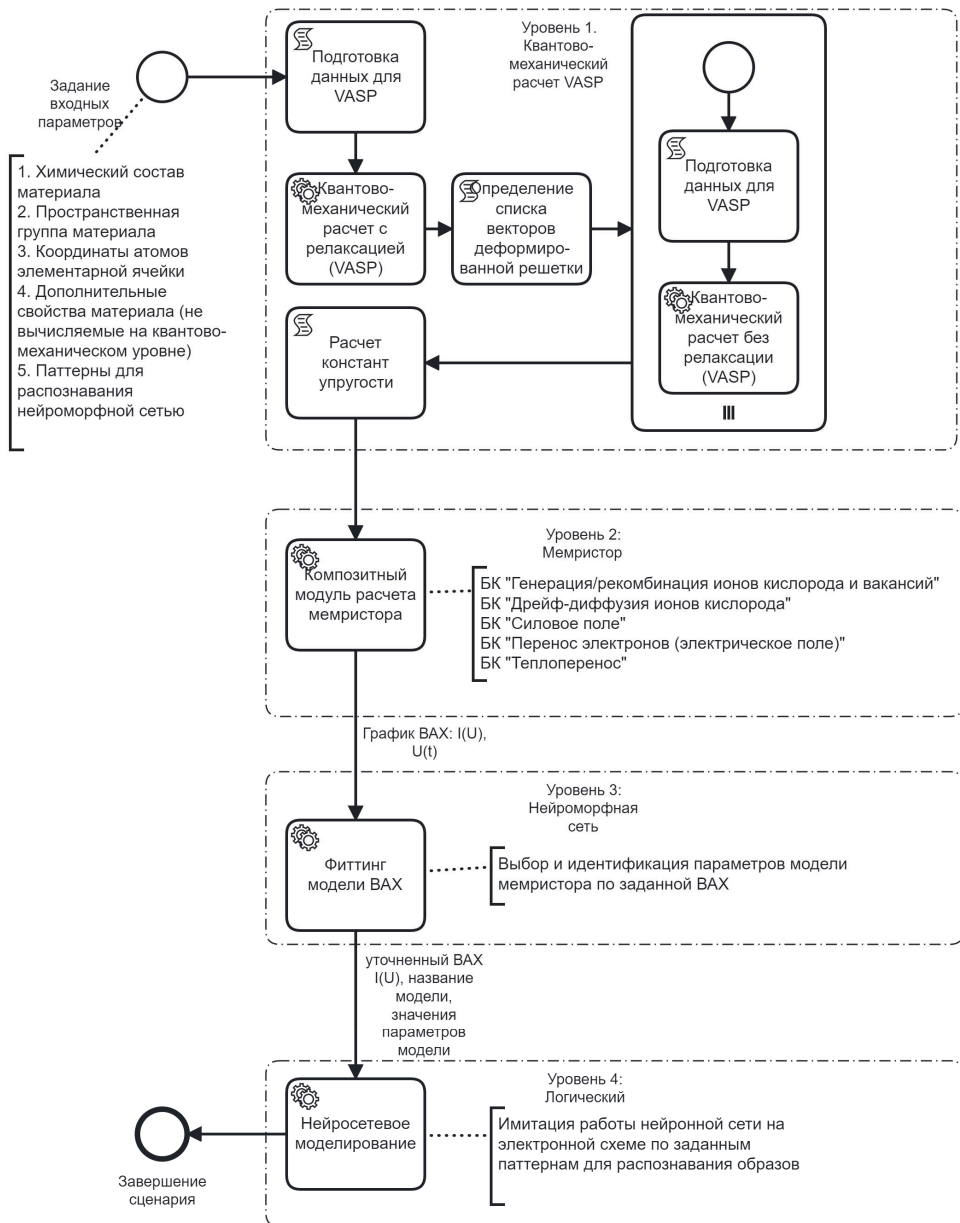


Рис 5.1. Диаграмма сценария нейроморфного моделирования

2. Квантово-механический расчет может производиться с помощью пакета VASP [67]. Чаще всего корректность расчета в квантовой механике зависит от множества специфических для пакета параметров, а особенно от выбранных псевдопотенциалов типов атомов. Корректность конкретного расчета также оценивает обычно специалист, что безусловно усложняет автоматизацию процесса в целом, но большое количество провалированных расчетов позволит создать необходимый массив знаний об использовании псевдопотенциалов, которые можно успешно использовать.

3. На втором масштабном уровне («мемристор») используется составной модуль расчета мемристора, в который интегрированы 5 физико-математических моделей. Их декомпозиция на отдельные модули в данном случае была признана нецелесообразной из-за интенсивного обмена данными в процессе расчетных итераций. Результатом работы модуля является вольтамперная характеристика (ВАХ) мемристора, смоделированная на выбранном материале.

4. Далее на базе вычисленной ВАХ производится фиттинг параметров модели мемристора, построенной на нем электронной схемы. На рисунке 5.2 показан пример входных параметров модуля фиттинга модели ВАХ мемристора.

5. На последнем масштабном уровне производится собственно симуляция работы нейронной сети, описанная в главе 4. Об успешности распознавания образов в симуляции делаются выводы о том, удачно ли подобран материал для исходного мемристора.

Введите параметры расчета

default

Химическая формула вещества

Химические элементы и их количества в элементарной ячейке

Hf

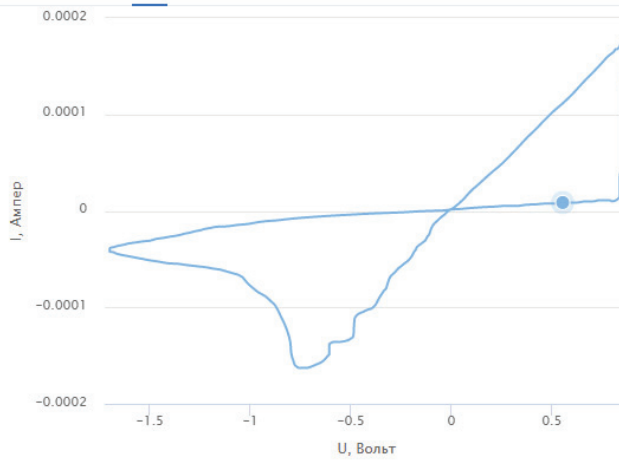
O

HfO₂

[Добавить хим. элемент](#)

Вольт-амперная характеристика

[Данные](#) [График](#)



[Upload File...](#)



Загрузите csv файл с тремя столбцами time,V,A

Загрузить данные из CIF файла

[Upload File...](#)



Drop file here

[Запустить](#)

Рис. 5.2. Параметры модуля фиттинга модели ВАХ мемристора

5.3. Выводы

При разработке платформы многомасштабного моделирования были определены роли пользователей платформы: ученый-исследователь, разработчик сценариев, разработчик расчетных модулей и системный разработчик, отвечающий за интеграцию модулей в платформу. Для каждой роли определен набор требований, методов и инструментов, позволяющий совместно достигать поставленные проектные цели.

Многомасштабное моделирование по своему определению требует привлечения ученых-экспертов из различных предметных областей, специализирующихся на отдельных физико-математических моделях, в свою очередь подразделяющихся на классы исследуемых объектов и их особенности. В связи с этим, поиск специалистов и организация проектных работ требует больших координационных усилий, временных и финансовых затрат. Предлагаемые методы автоматизации расчетных сценариев дают возможность стандартизировать план работ, привлекая специалистов определенных ролей для этапа создания многомасштабной модели и разработки и отладки соответствующих сценариев, а дальнейшие серии расчетов проводить автоматически с привлечением менее квалифицированных специалистов, так как использование готовых сценариев не требует глубоких знаний предметной области. База данных сценариев в этом случае представляет собой ценный актив экспертных знаний в виде исполняемых процедур. Таким образом, коллективы ученых-экспертов могут охватить больше актуальных областей исследований за фиксированные сроки, например, рассмотреть больше классов материалов или исследовать дополнительные свойства моделируемых материалов.

За счет автоматизации выполнения параметризованных расчетных сценариев сокращаются сроки выполнения многомасштабных расчетов при фиксированных ресурсах (вычислительных ресурсов и трудозатрат специалистов). В том числе, это достигается за счет автоматического

повторного использования результатов расчетов, хранящихся в базах данных сервисов расчетных модулей.

В этой главе показана применимость интеграционной платформы для проведения многомасштабных расчетов нейроморфных сетей от атомного уровня до макроуровня согласно сценариям, что позволяет автоматизировать процесс оптимизации их работы.

Список источников к главам 2, 3, 4, 5

1. Wong H.-S. P. et al. Metal–oxide RRAM // Proceedings of the IEEE, 2012. Vol. 100, no. 6, pp. 1951–1970. DOI: 10.1109/JPROC.2012.2190369.
2. Yang J. J., Strukov D. B., Stewart D. R. Memristive devices for computing // Nature Nanotechnology, 2013. Vol. 8, no. 1, pp. 13–24. DOI: 10.1038/nnano.2012.240.
3. Strukov D. B., Snider G. S., Stewart D. R., Williams R. S. The missing memristor found // Nature. Vol. 453, no. 7191, 2008, p. 80. DOI: 10.1038/nature06932.
4. Yang J. J., Pickett M. D., Xuema L., Ohlberg D. A. A., Stewart D. R., Williams R. S. Memristive switching mechanism for metal/oxide/metal nanodevices // Nature nanotechnology. Vol. 3, no. 7, 2008, p. 429. DOI: 10.1038/nnano.2008.160.
5. Pickett M. D., Stukov D. B., Borghetti J. L., Yang J. J., Snider G. S., Stewart D. R., Williams R. S. Switching dynamics in titanium dioxide memristive devices // Journal of Applied Physics. Vol. 106, no. 7, 2009, p. 074508. DOI: 10.1063/1.3236506.
6. Joglekar Y. N., Wolf S. J. The elusive memristor: properties of basic electrical circuits // European Journal of Physics. Vol. 30, no. 4, 2009, p. 661. DOI: 10.1088/0143-0807/30/4/001.
7. Biolek Z., Biolek D., Biolkova V. SPICE Model of Memristor with Nonlinear Dopant Drift // Radioengineering. Vol. 18, no. 2, 2009.
8. Prodromakis T., Peh B. P., Papavassiliou C., Toumazou C. A versatile memristor model with nonlinear dopant kinetics // IEEE Transactions on Electron Devices. Vol. 58, no. 9, 2011, pp. 3099–3105. DOI: 10.1109/TED.2011.2158004.

9. Zha J., Huang H., Liu Y. A novel window function for memristor model with application in programming analog circuits // IEEE Transactions on Circuits and Systems II: Express Briefs. Vol. 63, no. 5, 2015, pp. 423–427. DOI: 10.1109/TCSII.2015.2505959.
10. Kvatinsky S., Friedman E. G., Kolodny A., Weiser U. C. TEAM: Threshold adaptive memristor model // IEEE Transactions on Circuits and Systems I: Regular Papers. Vol. 60, no. 1, 2012, pp. 211–221. DOI: 10.1109/TCSI.2012.2215714.
11. Kvatinsky S., Ramadan M., Friedman E. G., Kolodny A. VTEAM: A general model for voltage-controlled memristors // IEEE Transactions on Circuits and Systems II: Express Briefs. Vol. 62, no. 8, 2015, pp. 786–790. DOI: 10.1109/TCSII.2015.2433536.
12. Yakopcic C., Taha T. M., Subramanyam G., Pino R. E., Rogers S. A memristor device model // IEEE electron device letters. Vol. 32, no. 10, 2011, pp. 1436–1438. DOI: 10.1109/LED.2011.2163292.
13. Rodriguez-Fernandez A., Cagli C., Perniola L., et al. Characterization of HfO₂-based devices with indication of second order memristor effects // Microelectronic Engineering. 2018. V. 195. P. 101-106. DOI:10.1016/j.mee.2018.04.006
14. Теплов Г. С., Горнев Е. С. Модель на языке verilog-а многоуровневого биполярного мемристора с учетом девиаций параметров переключения // Микроэлектроника. 2019. Т. 48, № 3. С. 163. DOI: 10.1134/S0544126919030104
15. Zheng G., Mohanty S. P., Koungianos E., Okobiah O. Polynomial Metamodel integrated Verilog-AMS for memristor-based mixed-signal system design // 2013 IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), Columbus, OH, 2013, pp. 916-919. DOI: 10.1109/MWSCAS.2013.6674799

16. Mladenov V. Analysis of Memory Matrices with HfO₂ Memristors in a PSpice Environment // *Electronics*, 8(4), 383, March 2019, p. 16. DOI: 10.3390/electronics8040383.
17. Agudov N. V., Safonov A. V., Krichigin A. V., Kharcheva A. A., Dubkov A. A., Valenti D., Guseinov D. V., Belov A. I., Mikhaylov A. N., Carollo A., Spagnolo B., Nonstationary distributions and relaxation times in a stochastic model of memristor // *Journal of Statistical Mechanics: Theory and Experiment*, 2020, 024003. DOI: 10.1088/1742-5468/ab684a
18. Mikhaylov A.N., Gryaznov E.G., Belov A.I., Korolev D.S., Sharapov A.N., Guseinov D.V., Tetelbaum D.I., Tikhov S.V., Malekhonova N.V., Bobrov A.I., Pavlov D.A., Gerasimova S.A., Kazantsev V.B., Agudov N.V., Dubkov A.A., Rosário C.M.M., Sobolev N.A., Spagnolo B. Field- and irradiation-induced phenomena in memristive nanomaterials // *Phys. Status Solidi C – Current Topics in Solid State Physics*, 2016;13(10-12):870-881 DOI: 10.1002/pssc.201600083
19. Filatov D.O., Vrzheshech D.V., Tabakov O.V., Novikov A.S., Belov A.I., Antonov I.N., Sharkov V. V., Koryazhkina M. N., Mikhaylov A. N., Gorshkov O. N., Dubkov A. A., Carollo A., Spagnolo B. Noise-induced resistive switching in a memristor based on ZrO₂(Y)/Ta₂O₅ stack // *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12), 124026. DOI: 10.1088/1742-5468/ab5704
20. Spagnolo B. , Valenti D., Guarcello C., Carollo A., Persano Adorno D., Spezia S., Pizzolato N., Di Paola B., Noise-induced effects in nonlinear relaxation of condensed matter systems // *Chaos, Solitons & Fractals*, 2015(81):412-424 DOI: 10.1016/j.chaos.2015.07.023
21. Spagnolo B., Guarcello C., Magazzù L., Carollo A., Persano Adorno D., Valenti D. Nonlinear Relaxation Phenomena in Metastable Condensed Matter Systems // *Entropy* 2017, 19(1), 20. DOI: 10.3390/e19010020

22. Васильев В. А., Чернов П. С. Математическое моделирование мемристора в присутствии шума // Матем. моделирование, 2014, Т. 26, № 1, с. 122–132
23. Морозов А.Ю., Ревизников Д.Л. Алгоритм адаптивной интерполяции на основе kd-дерева для численного интегрирования систем обыкновенных дифференциальных уравнений с интервальными начальными условиями // Дифференциальные уравнения. Т. 54. № 7. 2018. С. 963–974. DOI: 10.1134/S0374064118070130.
24. Fu C., Ren X., Yang Y.-F., Lu K., Qin W., Steady-state response analysis of cracked rotors with uncertain but bounded parameters using a polynomial surrogate method. // Commun. Nonlinear Sci. Numer. Simul. 2019, 68, 240–256, DOI:10.1016/j.cnsns.2018.08.004.
25. Fu C., Xu Y., Yang Y., Lu K., Gu F., Ball A., Response analysis of an accelerating unbalanced rotating system with both random and interval variables // J. Sound Vib. 2020, 466, 115047, DOI:10.1016/j.jsv.2019.115047.
26. Morozov A.Yu., Reviznikov D.L. Modeling of Dynamic Systems with Interval Parameters in the Presence of Singularities // Russian Journal of Nonlinear Dynamics. – 2020. – Vol. 16, No. 3.- P. 479–490. DOI: 10.20537/nd200306
27. Morozov A.Yu., Zhuravlev A.A., Reviznikov D.L. Sparse Grid Adaptive Interpolation in Problems of Modeling Dynamic Systems with Interval Parameters // Mathematics. – 2021. – Vol. 9, No. 4. – Article 298. DOI: 10.3390/math9040298
28. Смоляк С.А. Квадратурные и интерполяционные формулы на тензорных произведениях некоторых классов функций // Докл. АН СССР, 148:5 (1963), 1042–1045.

29. Judd K.L., Maliar S., Maliar L., Valero R., Smolyak Method for Solving Dynamic Economic Models: Lagrange Interpolation, Anisotropic Grid and Adaptive Domain // *Journal of Economic Dynamics and Control*. – 2014. – Vol. 44. – P. 92–123. DOI: 10.2139/ssrn.2306869
30. Bungartz H.-J. *Finite Elements of Higher Order on Sparse Grids*, Shaker, 1998. 127 p.
31. Морозов А.Ю., Ревизников Д.Л. Интервальный подход к решению задач параметрической идентификации динамических систем // *Дифференциальные уравнения*. 2022. Т. 58. № 7. С. 962–976. DOI: 10.31857/S0374064122070081.
32. Martyshov M.N., Emelyanov A.V., Demin, V.A. et al. Multifilamentary Character of Anticorrelated Capacitive and Resistive Switching in Memristive Structures Based on $(\text{Co-Fe-B})_x(\text{LiNbO}_3)_{100-x}$ Nanocomposite // *Phys. Rev. Applied*. 2020. V.14, № 3. 034016 DOI:10.1103/PhysRevApplied.14.034016
33. Rylkov V., Nikolaev S., Demin V. et al. Transport, Magnetic, and Memristive Properties of a Nanogranular $(\text{CoFeB})_x(\text{LiNbO}_y)_{100-x}$ Composite Material // *Journal of Experimental and Theoretical Physics*. 2018. V. 126. No. 3. pp.353-367. DOI:10.1134/S1063776118020152
34. Shary S.P. Randomized algorithms in interval global optimization // *Numerical Analysis and Applications*. – 2008. – Vol. 1, No. 4. – P. 376-389.
35. Chatterji B. N. // *IETE Journal of Education*. 1993. V. 34 № 1. P. 27-38. DOI: 10.1080/09747338.1993.11436397
36. Li C., Hu M., Li Y., Jiang H., Ge N., Montgomery E., Zhang J., Song W., Dávila N., Graves C. E., Li Z., Strachan J. P., Lin P., Wang Z., Barnell M., Wu Q., Williams R. S., Yang J. J., Xia Q. Analogue signal and image processing with large memristor crossbars // *Nature Electronics*. Vol. 1, no. 1, Dec. 2017, pp. 52–59. DOI:10.1038/s41928-017-0002-z.

37. Hu M., Graves C. E., Li C., Li Y., Ge N., Montgomery E., Davila N., Jiang H., Williams R. S., Yang J. J., Xia O., Strachan J. P. Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine // *Advanced Materials*. Vol. 30, no. 9, Jan. 2018, p. 1705914. DOI: 10.1002/adma.201705914.
38. Merolla P. A., Arthur J. V., Alvarez-Icaza R., et al. A million spiking-neuron integrated circuit with a scalable communication network and interface // *Science*. 2014. V. 345, № 6197. P. 668–673. DOI: 10.1126/science.1254642
39. Emelyanov A. V., Lapkin D. A., Demin V. A., et al. First steps towards the realization of a double layer perceptron based on organic memristive devices // *AIP advances*. 2016. V.6 № 11. P. 111301. DOI: 10.1063/1.4966257
40. Demin V.A., Nekhaev D.V., Surazhevsky I.A., et al. Necessary conditions for STDP-based pattern recognition learning in a memristive spiking neural network // *Neural Networks*. 2021. V. 134. P. 64–75. DOI: 10.1016/j.neunet.2020.11.005
41. Andreeva N.V., Ryndin E.A., Gerasimova M.I. Memristive logic design of multifunctional spiking neural network with unsupervised learning // *BioNanoScience*. 2020. V. 10. P. 824–833. DOI: 10.1007/s12668-020-00778-2
42. Emelyanov A. V., Nikiruy K. E., Serenko A. V., et al. Self-adaptive STDP-based learning of a spiking neuron with nanocomposite memristive weights // *Nanotechnology*. 2019. V. 31, № 4. P. 045201. DOI: 10.1088/1361-6528/ab4a6d
43. Diehl P., Cook M. Unsupervised learning of digit recognition using spike-timing-dependent plasticity // *Frontiers in Computational Neuroscience*, 2015. Vol. 9, art. 99, p. 9. DOI: 10.3389/fncom.2015.00099.
44. Ambrogio S. et al. Neuromorphic Learning and Recognition With One-Transistor-One-Resistor Synapses and Bistable Metal Oxide RRAM // *IEEE*

- Transactions on Electron Devices, 2016. Vol. 63, no. 4, pp. 1508–1515. DOI: 10.1109/TED.2016.2526647.
45. Guo Y., Wu H., Gao B., Qian H. Unsupervised Learning on Resistive Memory Array Based Spiking Neural Networks // *Front Neurosci*, 2019. Vol. 13, art. 812. DOI: 10.3389/fnins.2019.00812.
 46. Milo V. et al. Resistive switching synapses for unsupervised learning in feed-forward and recurrent neural networks // 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, 2018, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351824.
 47. Pedretti G., Bianchi S., Milo V., Calderoni A., Ramaswamy N., Ielmini D. Modeling-based design of brain-inspired spiking neural networks with RRAM learning synapses // 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2017, pp. 28.1.1–28.1.4. DOI: 10.1109/IEDM.2017.8268467.
 48. Milo V., Ielmini D., Chicca E. Attractor networks and associative memories with STDP learning in RRAM synapses // 2017 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2017, p. 11.2.1–11.2.4. DOI: 10.1109/IEDM.2017.8268369.
 49. Ielmini D., Ambrogio S., Milo V., Balatti S., Wang Z. Neuromorphic computing with hybrid memristive/CMOS synapses for real-time learning // 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, 2016, pp. 1386–1389. DOI: 10.1109/ISCAS.2016.7527508.
 50. Milo V. et al. Demonstration of hybrid CMOS/RRAM neural networks with spike time/rate-dependent plasticity // 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, 2016, pp. 16.8.1–16.8.4. DOI: 10.1109/IEDM.2016.7838435.
 51. Bhattacharya T., Parmar V., Suri M., MASTISK: Simulation Framework For Design Exploration Of Neuromorphic Hardware // 2018 International Joint

- Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1–9. DOI: 10.1109/IJCNN.2018.8489277.
52. Ambrogio S. et al. Novel RRAM-enabled 1T1R synapse capable of low-power STDP via burst-mode communication and real-time unsupervised machine learning // 2016 IEEE Symposium on VLSI Technology, Honolulu, HI, 2016, pp. 1–2. DOI: 10.1109/VLSIT.2016.7573432.
 53. Wenger C. et al. Inherent Stochastic Learning in CMOS-Integrated HfO₂ Arrays for Neuromorphic Computing // IEEE Electron Device Letters. Vol. 40, no. 4, April 2019, pp. 639–642. DOI: 10.1109/LED.2019.2900867.
 54. OpenMP // URL: <https://www.openmp.org/> (дата обращения: 02.04.2021)
 55. OpenMP technology // URL: <https://pvs-studio.com/ru/a/0057/> (date of access 02.04.2021)
 56. Фотохостинг Pinterest // URL: <https://ru.pinterest.com/pin/351912463120005/> (дата обращения: 02.09.2022).
 57. Gulli A., Pal S. Deep learning with Keras: implement neural networks with Keras on Theano and TensorFlow. Birmingham – Mumbai, UK: Packt Publishing Ltd, 2017, 490 p.
 58. MNIST CNN. Available at: https://keras.io/examples/mnist_cnn/ (accessed 01.10.2019)
 59. Train a simple deep CNN on the CIFAR10 small images dataset. Available at: https://keras.io/examples/cifar10_cnn/ (accessed 01.10.2019)
 60. Абгарян К. К. Многомасштабное моделирование в задачах структурного материаловедения. – М.: МАКС Пресс, 2017. 284 с.
 61. Абгарян К. К. Информационная технология построения многомасштабных моделей в задачах вычислительного материаловедения // Системы высокой доступности, 2018. Т. 15. № 2. С. 9-15.

62. Бродский Ю. И. Модельный синтез и модельно-ориентированное программирование. – М.: ВЦ РАН, 2013. 142 с.
63. Evans E. Domain-Driven Design: Tackling Complexity in the Heart of Software. – Addison Wesley Professional, 2003. 560 p.
64. Абгарян К. К., Гаврилов Е. С., Марасанов А. М. Информационная поддержка задач многомасштабного моделирования композиционных материалов // International Journal of Open Information Technologies, 2017. № 12. С. 24-29.
65. Martin Fowler, Pramod J. Sadalage. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence // Addison-Wesley Professional, 2012, 190 с.
66. Программный пакет для молекулярно-динамических и других расчетов Materials Studio [Электронный ресурс] URL: <https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-materials-studio/>
67. Программный пакет квантово-механических расчетов VASP [Электронный ресурс] URL: <https://www.vasp.at>
68. Программный пакет квантово-механических расчетов QuantumEspresso [Электронный ресурс] URL: <https://www.quantum-espresso.org>
69. W. Kohn, L.J. Sham. Self – consistent equations including exchange and correlation effects. – Phys. Rev. 140:A. 1965. P. 1133-1138.
70. Gao D., Deng H., Heinisch H., Kurtz R. A new Fe-He interatomic potential based on abinitio calculations in α -Fe. J. Nucl. Mater. 418, 115. 2011.
71. Byeong-Joo Lee, M.I. Baskes, Hanchul Kim, and Yang Koo Cho. Second nearest-neighbor modified embedded atom method potentials for bcc transition metals //Phys. Rev. B, V. 64, p.184102. 2001. DOI: 10.1103/PhysRevB.64.184102

72. Sadi, T., Mehonic, A., Montesi, L., Buckwell, M., Kenyon, A. and Asenov, A. Investigation of resistance switching in SiO_x RRAM cells using a 3D multi-scale kinetic Monte Carlo simulator. *Journal of Physics: Condensed Matter*, 30(8), 084005. 2018. DOI: 10.1088/1361-648X/aaa7C1
73. Mordkovich, V. N., Abgaryan, K. K., Reviznikov, D. L. and Leonov, A. V. Simulation of Hall Field Elements Based on Nanosized Silicon-on-Insulator Heterostructures // ISSN 1063-7397, *Russian Microelectronics*, 2021, Vol. 50, No. 8, pp. 1–6.
74. Спецификация BPMN [Электронный ресурс] URL: <https://www.bpmn.org>
75. Машинный протокол и формат файлов для обмена кристаллографической информацией CIF [Электронный ресурс] URL: <https://www.iucr.org/resources/cif>
76. Открытая база данных кристаллических структур [Электронный ресурс] URL: <http://crystallography.net/cod/>

*The work was supported by the Russian Foundation for Basic Research
under grant No.19-29-03051 mk*

K.K. Abgaryan, D.L. Reviznikov, A.A. Zhuravlev, A.Yu. Morozov, E.S. Gavrilov

Multiscale modeling of neuromorphic systems

Multiscale approaches, models and algorithms for designing neuromorphic devices for the memory of new generation computers are presented. The developed approaches make it possible to solve problems associated with simulating modeling of the work of neuromorphic networks in intellectual analysis modes of data and machine learning.

The issues related to the construction of a computing model of formation/destruction of conductive channels in memristive elements underlying neuromorphic networks are considered. New algorithms for modeling the work of the neuromorphic network have been created, taking into account stochastic effects, as well as original methods and means of simulating modeling of teaching a neuromorphic network.

The main approaches are presented when creating software for simulating modeling of neuromorphic networks based on the integration platform of multiscale modeling, combining information flows at various large -scale levels, including the level of a new resistive memory element, the level of neuromorphic network and the level of simulation of the neuromorphic network on precedents .

The book is intended for scientists, specialists in the field of computing electronics, senior students and graduate students of technical universities.

Key words: mathematical modeling, neuromorphic systems, analog neural networks, memristive elements, stochastic dynamics, filament, multiscale modeling, hybrid software systems, high performance systems

Научное издание

АБГАРЯН Каринэ Карленовна
РЕВИЗНИКОВ Дмитрий Леонидович,
ЖУРАВЛЕВ Андрей Андреевич,
МОРОЗОВ Александр Юрьевич,
ГАВРИЛОВ Евгений Сергеевич

МНОГОМАСШТАБНОЕ МОДЕЛИРОВАНИЕ
НЕЙРОМОРФНЫХ СИСТЕМ

Монография

Издательство «МАКС Пресс»
Главный редактор: *Е.М. Бугачева*
Обложка: *А.В. Кононова*

Отпечатано с готового оригинал-макета

Подписано в печать 21.12.2022 г.
Формат 70х100 1/16. Усл.печ.л. 10,4.
Тираж 500 (1–150) экз. Изд. № 198

Издательство ООО «МАКС Пресс»
Лицензия ИД N 00510 от 01.12.99 г.

119992, ГСП-2, Москва, Ленинские горы,
МГУ им. М.В. Ломоносова, 2-й учебный корпус, 527 к.
Тел. 8(495)939-3890/91. Тел./Факс 8(495)939-3891.

Отпечатано в полном соответствии с качеством
предоставленных материалов в ООО «Фотоэксперт»
109316, г. Москва, Волгоградский проспект, д. 42,
корп. 5, эт. 1, пом. I, ком. 6.3-23Н

