_ НАНОЭЛЕКТРОНИКА И НЕЙРОМОРФНЫЕ ____ ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

УДК 537.9+004.8

МАССИВЫ НАНОКОМПОЗИТНЫХ КРОССБАР-МЕМРИСТОРОВ ДЛЯ РЕАЛИЗАЦИИ ФОРМАЛЬНЫХ И ИМПУЛЬСНЫХ НЕЙРОМОРФНЫХ СИСТЕМ

© 2022 г. А. И. Ильясов^{1,2,*}, К. Э. Никируй¹, А. В. Емельянов^{1,**}, К. Ю. Черноглазов¹, А. В. Ситников^{1,3}, В. В. Рыльков^{1,4}, В. А. Демин¹

¹Национальный исследовательский центр "Курчатовский институт", Москва, Россия

²Московский государственный университет им. М.В. Ломоносова, Москва, Россия

³Воронежский государственный технический университет, Воронеж, Россия

⁴Фрязинский филиал Института радиотехники и электроники им. В.А. Котельникова РАН, Фрязино, Россия

*E-mail: sashailyasov99@gmail.com **E-mail: emelyanov_av@nrcki.ru Поступила в редакцию 29.11.2021 г. После доработки 29.11.2021 г. Принята к публикации 06.12.2021 г.

Мемристоры — многообещающие устройства для использования в качестве синапсов в аппаратных искусственных формальных и спайковых нейронных сетях. Их применение может увеличить энергоэффективность таких сетей при решении различных когнитивных задач. При этом для реализации аппаратных нейронных сетей требуется большое количество достаточно плотно упакованных структур, чего можно добиться, создавая мемристивные матрицы в кроссбар-архитектуре. Разработаны и изготовлены кроссбар-массивы нанокомпозитных мемристоров на кремниевых подложках с защитой от краевых эффектов, исследованы их резистивные переключения, создана формальная нейронная сеть на их основе, способная распознавать простые образы после загрузки предобученной карты весов, а также продемонстрирована возможность точной настройки проводимости мемристоров кроссбара различными способами, применимыми при создании более сложных спайковых и формальных нейронных сетей.

DOI: 10.56304/S1992722322010058

введение

В настоящее время вычислительные системы, построенные по классической архитектуре фон Неймана, сталкиваются с рядом трудностей, связанных со свойственными ей неотъемлемыми недостатками и ограничениями. В первую очередь это повышенное энергопотребление и ограничение скорости работы из-за необходимости постоянного проведения операций чтения из памяти/записи в память данных и результатов вычислений [1, 2]. Поэтому реализация формальных нейронных сетей (ФНС) на устройствах классической полупроводниковой электроники становится все более энергозатратной с увеличением их размера.

Одним из многообещающих путей решения этой проблемы является реализация нейронных сетей на базе мемристоров. Мемристоры – устройства, способные изменять свою проводимость под действием приложенного напряжения, превышающего некоторый порог, а также сохранять резистивное состояние после прекращения

внешнего воздействия [3-7]. Это свойство делает их прекрасными кандидатами для реализации нейронных сетей – как формальных, так и спайковых (импульсных). В последнее время были продемонстрированы, например, перцептроны [8, 9], кратковременная-долговременная память [10], резервуарные вычислители [11], а также распознаватели простых образов [12] и обучение шумовым сигналом [13] в спайковых нейронных сетях. Такие сети потенциально являются еще более энергоэффективными, чем формальные, однако эффективные алгоритмы обучения для них находятся в разработке [14]. Поэтому построение ФНС на базе мемристоров в данный момент представляется удачным компромиссом между потенциальной эффективностью импульсных сетей и практической возможностью аппаратной реализации ФНС.

Такой подход, однако, имеет свои сложности, главная из которых — необходимость точно знать проводимости каждого мемристора и иметь возможность их тонкой настройки в процессе обучения, что может быть затруднительно из-за разброса характеристик мемристоров от устройства к устройству [15]. Для решения этой проблемы может быть использован следующий подход: первоначальная карта проводимостей (весов связей) рассчитывается отдельно (*ex situ*) и только потом загружается в массив мемристоров. После этого определяется эффективность работы сети и при необходимости ее "дообучают" *in situ*, подстраивая начальную карту весов для обеспечения наибольшей точности работы сети.

Для разработки нейроморфных вычислительных систем (НВС), нацеленных на решение сложных когнитивных задач, необходимо использовать достаточно большие массивы синаптических мемристивных элементов с высокой плотностью размещения на "кристалле", достигаемой в топологии кроссбар (элементы находятся на пересечениях взаимно перпендикулярных электродных шин "битов" и "слов") [14]. В этой ситуации одними из важных требований, предъявляемых к мемристивным элементам, являются малая энергия потребления при резистивном переключении (РП) элемента и низкие напряжения РП для эффективной интеграции матриц с современными процессорными КМОП-устройствами (комплементарная логика на транзисторах металл-оксид-полупроводник), работающими при напряжениях не более 3 В. При этом важным параметром мемристора является многоуровневый характер РП (пластичность), позволяющий задавать множество промежуточных состояний веса связи между максимальным и минимальным.

Настоящая работа посвящена построению простой ФНС на базе массива нанокомпозитных мемристоров, объединенных в кроссбар-архитектуру. Созданная сеть обучается распознаванию образов "10" и "01". Также демонстрируются кривые пластичности, зависящей от времени прихода импульсов (Spike-timing dependent plasticity, **STDP**) [19] мемристоров кроссбара, что позволяет сделать вывод о возможности его использования и для реализации спайковых HBC.

ОБРАЗЦЫ И МЕТОДЫ

В [16–18] было показано, что перспективным направлением в создании кроссбар-массивов низковольтных мемристоров с малым энергопотреблением может быть их разработка с использованием двухслойных структур металл/нанокомпозит/диэлектрик/металл (**M/HK/Д/M**) на основе нанометровых слоев HK = $(Co_{40}Fe_{40}B_{20})_x(LiNbO_3)_{100-x}$ и Д = = LiNbO₃ (**LNO**) с суммарной толщиной активной области около 100 нм. Один из проблемных факторов в создании кроссбар-структур при относительно толстых шинах (≥100 нм) связан с влиянием краевых эффектов, приводящих к токовым утечкам и локальному электрическому пробою структуры. Это требует изоляции рабочей области нанокомпозита по краям структуры от верхних металлических шин, которая может быть создана с использованием плазмохимического осаждения (например, Si_3N_4) при температуре 300°С. В этом случае формируется плотный диэлектрик. При относительно небольшой толщине (5–10 нм) диэлектрической пленки эффекты прямого туннелирования электронов пренебрежимо малы.

С учетом этого для создания мемристивных матриц, размещенных на SiO₂/Si-подложках диаметром 51 мм, был модифицирован соответствующий технологический маршрут, основанный на использовании лазерной фотолитографии (литограф Heidelberg 66fs; разрешение 1 мкм), плазмохимического осаждения (Trion Oracle III) и ионно-лучевого распыления составной мишени на оригинальной установке, позволяющей в едином цикле получать слои НК различного состава [20]. Важная особенность при создании ансамбля M/HK/LNO/M-структур заключалась в использовании при осаждении НК- и LNO-слоев теневых масок, что позволило сформировать НК-полоски необходимых размеров и при этом избежать использования химического травления, которое трудно осуществимо при различных составах НК. На данном этапе использовали плазмохимическое осаждение для получения как нижнего, так и верхнего защитных слоев (Si_3N_4) . При этом чтобы подавить нежелательное влияние краевых эффектов, приводящих к токовым утечкам и локальному электрическому пробою структуры, размер окон (47×47 мкм²) в защитном слое диэлектрика Si₃N₄ на обеих шинах был выбран заметно меньшим, чем их ширина (50 мкм) (рис. 1). Были изготовлены кроссбар-структуры с числом элементов 4 \times 4 площадью $S = 50 \times 50$ мкм². Содержание металла в исследуемом кроссбаре составляло $x \approx 19$ ат. %.

Исследования мемристивных свойств М/HK/LNO/М кроссбар-элементов проводили при комнатной температуре с помощью многофункционального источника-измерителя NI PXIe-4140 (National Instruments) и аналитической зондовой станции PM5 (Cascade Microtech), снабженной оптической системой микрометрического перемещения зондов PSM-100 (Motic).

Эксперименты по реализации формальной сети с использованием исследованных мемристоров проводили в несколько этапов. На первом этапе карту весов переносили в кроссбар-матрицу путем перевода мемристоров в состояния с соответствующими весам проводимостями. Для этого контакт строки, на которой находится мемристор, подключали к положительному выходу одного из каналов источника-измерителя,



Рис. 1. Схематическое изображение сечения отдельного мемристивного кроссбар-элемента на базе структур $M/(CoFeB)_x(LiNbO_3)_{100-x}/LiNbO_3/M$ с защитными изолирующими слоями Si₃N₄, выращенными плазмо-химическим осаждением

контакт соответствующего столбца — к отрицательному выходу ("земле"). Установку в нужное состояние проводили по специальному алгоритму записи—чтения [21] с помощью последовательности импульсов различной длительности с точностью до 5% от целевого значения сопротивления.

После переноса всей карты (настройки проводимостей всех четырех мемристоров) исследовали возможность классификации простых образов "предобученной" сетью. Для этого положительные выходы двух каналов источника-измерителя подключали к двум строковым контактам кроссбар-матрицы, а контакты 2-го и 3-го ее столбцов через пару резисторов номиналом $R_0 = 10$ Ом подключали к отрицательным контактам (общей земле) каналов источника-измерителя. Исследования работы ФНС проводили с помощью четырехканального осциллографа WaveSurfer 4000HD (Teledyne LeCroy).

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

На основе кроссбара НК-мемристоров была построена HBC, способная к обучению решению задачи классификации простых образов. Обучение в такой сети проводилось *ex situ*. После предварительного обучения сети на внешнем компьютере полученные веса должны быть записаны в кроссбар-матрицу мемристоров. После этого исследуется возможность классификации простых образов полученной сетью. В случае неудовлетворительной точности классификации сеть может быть дообучена *in situ* путем более тонкой подстройки весов мемристоров. В качестве мемристивной матрицы был использован участок 2×2 , расположенный в центре кроссбара 4×4 . Схема эксперимента представлена на рис. 2. На строковые контакты кроссбара подавались образы "01" и "10" в виде импульсов напряжения. Выходной сигнал снимался путем считывания токов с контактов столбцов кроссбара.

На рис. 3 представлены вольт-амперные характеристики (**BAX**) выбранных мемристоров. Скорость развертки по напряжению составляла 2 В/с, тогда как токи ограничения были установлены на уровне 100 мА для обеих полярностей приложенного напряжения. Для каждого мемристора снято по 50 кривых ВАХ. Отметим, что их различия от цикла к циклу невелики, в то время как различия от устройства к устройству более заметны, но все же некритичны для выбранной задачи классификации. Механизм РП в НК-мемристорах основан на движении вакансий кислорода под действием электрического поля по большому числу филаментов [18].

По данным ВАХ рассчитаны "окна" сопротивлений мемристоров (значения $R_{\rm on}$ и $R_{\rm off}$). Это необходимо для перевода весов связей сети в реальные проводимости. Сопротивления $R_{\rm on}$ и $R_{\rm off}$ рассчитывали на уровне напряжения чтения $U_{\rm read} = 0.3$ В после положительной и отрицательной части кривой ВАХ соответственно. На рис. 4 представлены зависимости сопротивлений $R_{\rm on}$ и $R_{\rm off}$ мемристоров от номера цикла ВАХ (на рис. 2–4 одинаковые цвета соответствуют одним и тем же мемристивным элементам в кроссбаре). Низкоомные (on) и высокоомные (off) состояния разделены довольно заметной маржинальной областью с центром в области 100 Ом. Видно, что соотноше-



Рис. 2. Схема эксперимента по реализации формальной сети на базе кроссбар-мемристоров.



Рис. 3. Вольт-амперные кривые использованных мемристоров.

ние сопротивлений R_{on}/R_{off} имеет значение не менее 1:3 для всех четырех мемристоров и всех циклов снятия ВАХ. Отметим, что при использовании напряжений с большими амплитудами возможно увеличение соотношения сопротивлений не менее чем до 1:10.

В качестве модельных образов в задаче классификации формальной сетью на основе мемристоров были выбраны следующие комбинации: "01" (напряжение логического нуля на второй строке кроссбара и логической единицы на третьей) и "10" (рис. 2). Длительность предъявления образа составляла 10 мс, напряжение логической единицы $U_{\text{high}} = 0.3$ В.

Исследовали работу сети с двумя картами весов. В первом варианте все веса были равны ($G_1 = 10 \text{ мCm}$), сеть не обучена. Результат этого случая отражен на рис. 5. Можно заметить, что при подаче каждого из образов выходные токи практически неотличимы, следовательно, точность



Рис. 4. Сопротивления использованных мемристоров, полученные путем обработки ВАХ, представленных на рис. 3.



Рис. 5. Входные и выходные сигналы необученной сети. На верхней панели показаны выходные токи сети, горизонтальная линия соответствует пороговому значению тока *I*_{threshold} = 4 мкА, выбранному для логического разделения разных входных паттернов; на нижней – входные напряжения сети, которые подавались на контакты строк и соответствовали образам "01" и "10".

классификации в данном случае оказывается нулевой.

После этого была применена процедура *ex situ*-обучения сети. В ней высокие веса располагались на побочной диагонали квадрата 2 × 2, а низкие — на главной. Тепловая карта весов изображена на рис. 6. Выбор "высокопроводящей" диа-

гонали обусловлен следующим соображением: ненулевое сопротивление проводящих шин в кроссбаре может негативно влиять на точность распознавания образов, уменьшая окно сопротивлений каждого из мемристоров, а также смещая эти окна в общем случае на разные величины, зависящие от суммарного расстояния от



Рис. 6. Тепловая карта проводимостей мемристоров, соответствующая карте весов связей в предобученной сети.

мемристора до контактов, соответствующих его строке и столбцу. Несмотря на то что контактные площадки есть на всех четырех сторонах кроссбара, во всех экспериментах использовались контакты слева (для строк) и сверху (для столбцов). При этом для мемристоров, расположенных слева снизу и справа сверху от центрального квадрата 2×2 суммарные расстояния до этих контактов, а значит, и добавочные сопротивления оказываются примерно одинаковыми. Именно поэтому данные мемристоры были выбраны в качестве высокопроводящих.

После загрузки карты весов в кроссбар (перевода всех мемристоров в соответствующие состояния проводимости) был повторен эксперимент по распознаванию простых образов. Исходя из карты проводимостей при правильной работе сети ожидалось, что для образа "01" выходной ток с контакта второго столбца будет больше, чем с контакта третьего; и наоборот, для образа "10". Результаты эксперимента показаны на рис. 7. Видно, что с новой картой весов ток на выходе, соответствующем предъявляемому в данный момент образу, превышает ток на другом выходе более чем в 3 раза, уровень $I_{\rm threshold}$ четко разделяет эти два значения тока. Предобученная сеть стала распознавать образы так, как и ожидалось.

Несмотря на то что в данном случае дообучения сети после загрузки начальной карты весов не потребовалось, для больших размеров HBC и/или паттернов, кодируемых промежуточными значениями в диапазоне (0, 1), оно может быть необходимо. При этом целевое изменение проводимости ΔG каждого мемристора (веса связи) должно рассчитываться по формулам метода обратного распространения ошибки исходя из ошибки распознавания на данной итерации [8, 9, 22].

Помимо расчета значения ΔG нужно иметь алгоритм реального изменения на данную величину проводимости каждого отдельного мемристора. Наиболее универсальным и простым с точки зрения схемотехники является метод приведения мемристора в нужное состояние при помощи двух видов импульсов с заданными параметрами: положительного (потенциирующего), повышающего проводимость, и отрицательного (депрессирующего), уменьшающего ее. Для реализации этого метода нужно знать, как выглядит кривая зависимости проводимости мемристора от числа поданных импульсов, а также выбрать их параметры такими, чтобы эта кривая была максимально близкой к прямой линии.

Изученные мемристоры демонстрируют довольно резкое изменение проводимости первыми несколькими импульсами при смене их полярности. Особенно сильно это выражается, когда они начинают потенцироваться ($\Delta G > 0$) положительными импульсами из состояния, близкого к R_{off} . В этом случае каждый импульс может изменять их сопротивление на десятки процентов от всего окна проводимости. Поэтому для данных мемристоров было решено использовать следующий алгоритм: если по результатам итерации требуется



Рис. 7. Входные и выходные сигналы предобученной сети. Обозначения те же, что на рис. 5.

потенциация мемристора, то он вначале должен приводиться в состояние с заведомо большей проводимостью, чем целевая (например, одним импульсом, переводящим в низкоомное состояние), а затем серией импульсов отрицательной полярности происходит точная настройка проводимости.

Для НК-мемристоров были выбраны следующие оптимальные параметры импульсов: амплитуды -2.5 и +2.8 В, одинаковые длительности в 30 мс и прямоугольная форма. При таком выборе параметров положительные импульсы практически сразу переводят мемристор в низкоомное состояние, а отрицательные импульсы меняют проводимость достаточно плавно (рис. 8). Зависимость этого изменения от числа импульсов определяли следующим образом: сначала мемристор приводили в высокопроводящее состояние положительными импульсами, затем на него подавали 75 отрицательных, причем после каждого импульса проводили считывание текущей проводимости мемристора. Такой измерительный цикл повторяли 15 раз для набора статистики. Подобное плавное изменение проводимости НК-мемристоров одинаковыми импульсами при одной полярности удобно с практической точки зрения схемотехнической реализации ядра НВС в будущем интегральном исполнении.

Также была изучена возможность изменения проводимости мемристоров по локальным биоподобным правилам типа STDP. Для этого на отдельный мемристор с источника-измерителя подавали пары прямоугольных биполярных импульсов с амплитудами и длительностями плеч, равными 1.2 В и 30 мс соответственно, наложенных друг на друга с различной задержкой по времени. Перед этим мемристор приводили в среднее значение проводимости *G*_{initial} = 10 мСм, а по-



Рис. 8. Кривые изменения проводимости мемристора при последовательной подаче одинаковых депрессирующих импульсов в зависимости от числа импульсов. Серыми линиями обозначены результаты каждого цикла измерения, точками – значения проводимости после *i*-го импульса, усредненные по 15 циклам. Точки соединены жирной линией для наглядности.



Рис. 9. STDP-кривые четырех исследованных мемристоров.

сле подачи проводимость измерялась снова. Ее относительное изменение в зависимости от временного промежутка между импульсами для четырех различных мемристоров изображена на рис. 9. Каждая точка на графиках соответствует медианному усреднению по пяти повторениям подачи импульсов с одинаковой задержкой. Данные результаты свидетельствуют о возможности изменения проводимости мемристоров кроссбара по правилам STDP, что может быть использовано при построении спайковых HBC на их основе.

ЗАКЛЮЧЕНИЕ

Сформированы массивы мемристивных кроссбар-элементов на основе нанокомпозита CoFeB-LiNbO₃ с прослойкой чистого LiNbO₃ на кремниевой подложке, демонстрирующие обратимые РП с окном сопротивлений $R_{\rm off}/R_{\rm on}$ порядка 10 и хорошей повторяемостью от цикла к циклу и от устройства к устройству. На базе разработанных массивов мемристоров продемонстрирована формальная сеть 2 × 2, способная распознавать простые образы после загрузки предобученной матрицы весов. Продемонстрирована возможность изменения проводимости мемристоров в кроссбар-архитектуре по локальным биоподобным правилам STDP, а также прецизионного приведения их в заранее заданное состояние при помощи последовательности одинаковых импульсов, что может быть использовано при построении более сложных спайковых и формальных НВС соответственно.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследова-

ний (грант № 18-29-23041) в части импульсного обучения мемристоров и Российского научного фонда (грант № 18-79-10253) в части построения и обучения формальной нейроморфной сети. Исследования выполнены на оборудовании Ресурсного центра электрофизических методов Национального исследовательского центра "Курчатовский институт" (№ 2753).

СПИСОК ЛИТЕРАТУРЫ

- Jouppi N.P., Young C., Patil N. et al. // Proc. Int. Symp. Comput. Archit. 2017. P. 1. https://doi.org/10.1145/3079856.3080246
- 2. *Hubara I., Courbariaux M., Soudry D. et al.* // J. Mach. Learn. Res. 2018. V. 18. P. 1.
- 3. Strukov D.B., Snider G.S., Stewart D.R., Williams R.S. // Nature. 2008. V. 453. № 7191. P. 80. https://doi.org/10.1038/nature06932
- Mikhaylov A.N., Guseinov D.V., Belov A.I. et al. // Chaos, Solitons Fractals. 2021. V. 144. P. 110723. https://doi.org/10.1016/j.chaos.2021.110723
- Mikheev V., Chouprik A., Lebedinskii Y. et al. // Nanotechnology. 2020. V. 31. № 21. P. 215205. https://doi.org/10.1088/1361-6528/ab746d
- Bobylev A.N., Udovichenko S.Y. // Russ. Microelectron. 2016. V. 45. № 6. P. 396. https://doi.org/10.1134/S1063739716060020
- Wang W., Panin G.N., Fu X. et al. // Sci. Rep. 2016. V. 6. № 1. P. 31224. https://doi.org/10.1038/srep31224
- Prezioso M., Merrikh-Bayat F., Hoskins B.D. et al. // Nature. 2015. V. 521. № 7550. P. 61. https://doi.org/10.1038/nature14441
- 9. Emelyanov A.V., Lapkin D.A., Demin V.A. et al. // AIP Adv. 2016. V. 6. № 11. https://doi.org/10.1063/1.4966257
- Li C., Wang Z., Rao M. et al. // Nat. Mach. Intell. 2019.
 V. 1. № 1. P. 49. https://doi.org/10.1038/s42256-018-0001-4
- Moon J., Ma W., Shin J.H. et al. // Nat. Electron. 2019.
 V. 2. № 10. P. 480. https://doi.org/10.1038/s41928-019-0313-3
- Feldmann J., Youngblood N., Wright C.D. et al. // Nature. 2019. V. 569. № 7755. P. 208. https://doi.org/10.1038/s41586-019-1157-8
- 13. Surazhevsky I.A., Demin V.A., Ilyasov A.I. et al. // Chaos, Solitons Fractals. 2021. V. 146. P. 110890. https://doi.org/10.1016/j.chaos.2021.110890
- 14. Xia Q., Yang J.J. // Nat. Mater. 2019. V. 18. № 4. P. 309. https://doi.org/10.1038/s41563-019-0291-x
- Ryu J.-H., Kim S. // Chaos, Solitons Fractals. 2020.
 V. 140. P. 110236. https://doi.org/10.1016/j.chaos.2020.110236

- Рыльков В.В., Емельянов А.В., Николаев С.Н. и др. // ЖЭТФ. 2020. Т. 158. № 1 (7). С. 164. https://doi.org/10.31857/s0044451020070159
- 17. Emelyanov A.V., Nikiruy K.E., Serenko A.V. et al. // Nanotechnology. 2020. V. 31. № 4. P. 045201. https://doi.org/10.1088/1361-6528/ab4a6d
- Martyshov M.N., Emelyanov A.V., Demin V.A. et al. // Phys. Rev. Appl. 2020. V. 14. № 3. P. 034016. https://doi.org/10.1103/PhysRevApplied.14.034016
- 19. *Querlioz D., Bichler O., Dollfus P., Gamrat C. //* IEEE Trans. Nanotechnol. 2013. V. 12. № 3. P. 288. https://doi.org/10.1109/TNANO.2013.2250995
- 20. Никируй К.Э., Емельянов А.В., Рыльков В.В. и др. // Радиотехника и электроника. 2019. Т. 64. № 10. С. 1019. https://doi.org/10.1134/S0033849419100103
- Никируй К.Э., Емельянов А.В., Демин В.А. и др. // Письма в ЖТФ. 2018. Т. 44. № 10. С. 20. https://doi.org/10.21883/PJTF.2018.10.46095.17099
- 22. Mikhaylov A.N., Pigareva Y.I., Pimashkin A.S. et al. // IEEE Trans. Emerg. Top. Comput. Intell. 2018. V. 2. № 5. P. 371. https://doi.org/10.1109/TETCI.2018.2829922