

УДК 004.383.8.032.26

Моделирование процесса функционирования распределенной системы обработки информации на базе мемристорных нейронных сетей

Никишов Д.А., Антонов А.М., Жиганова Е.А., Жиганов С.Н., Борданов И.А., Данилин С.Н., Щаников С.А.

Статья посвящена решению актуальной научно-практической задачи, связанной с разработкой методов и программно-аппаратных средств моделирования, распределенных системы обработки информации на базе мемристорных нейронных сетей. В работе приведен оригинальный подход к моделированию таких систем, основанный на интеграции схемотехнических SPICE моделей, эмулирующих работу аналоговых частей вычислителя в LTSpice на физическом уровне, и моделей на языке программирования Python, эмулирующих работу устройств управления на системном уровне.

Ключевые слова: искусственные нейронные сети, мемристоры, распределенная обработка данных, аппаратные ускорители искусственного интеллекта.

Введение

В настоящее время одним из наиболее перспективных направлений исследований в области создания распределенных вычислительных систем, предназначенных для аппаратного ускорения работы алгоритмов искусственного интеллекта (ИИ), является разработка мемристивных электронных компонентов и устройств на их основе [1]. Такие устройства, часто называемые аппаратный ускоритель ИИ (от англ. «AI accelerator»), позволяют выполнять операции матрично-векторного умножения в аналоговом виде за счет естественных законов Ома и Кирхгофа, что позволяет повысить скорость обработки информации и снизить энергопотребление [2]. Важной задачей в данном научном направлении является разработка методов и программных средств для моделирования таких систем [3].

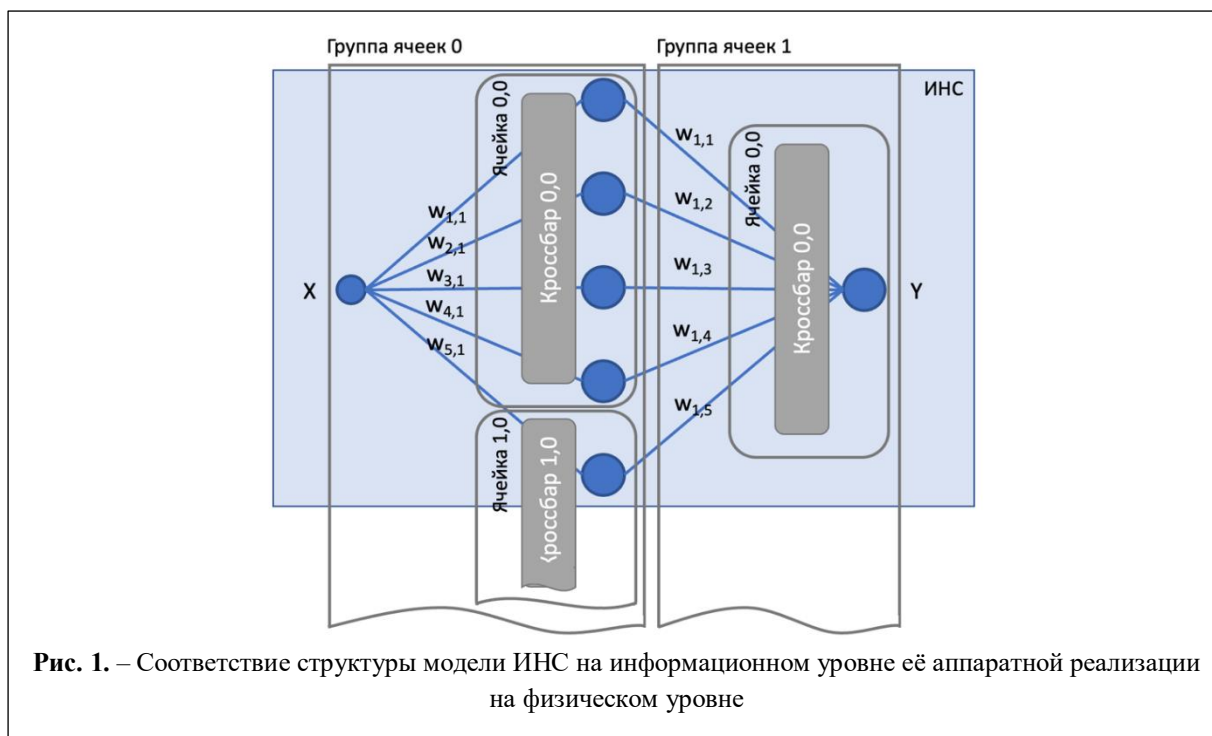
Авторами разрабатывается программный комплекс (ПК) «МемриСим» [4], который можно использовать для проведения численных экспериментов и исследования компьютерных моделей искусственных нейронных сетей (ИНС) на базе мемристивных устройств. ПК «МемриСим» является модульным, кроссплатформенным, расширяемым программным обеспечением, написанным на языке программирования Python и имеющим интерфейсы взаимодействия с программными моделями ИНС, созданными в Keras или NeuroLab. В данной статье описан

новый модуль, позволяющий подготовить модель ИНС к выполнению на аппаратном ускорителе ИИ на базе мемристивных устройств и провести программную симуляцию его работы.

1. Архитектура аппаратного ускорителя

Аппаратный ускоритель ИИ на базе мемристоров имеет следующую архитектуру. Базовым вычислительным узлом ускорителя, выполняющим матрично-векторное умножение (МВУ) является ячейка (рис. 1). Ячейка содержит кроссбары, количество которых можно масштабировать по горизонтали и вертикали, и устройство управления на базе ПЛИС. Ячейки можно сконфигурировать в группы ячеек, соответствующие слоям ИНС. Обработка данных внутри группы происходит параллельно, а между группами – последовательно. Группы ячеек в целом составляют блок обработки данных.

Такой блок получает на вход данные, «пропускает» их через модель ИНС и выдает результат. На самом верхнем уровне масштабирования системы блоки обработки данных можно объединить в вычислительный кластер (рис. 2). Такой кластер



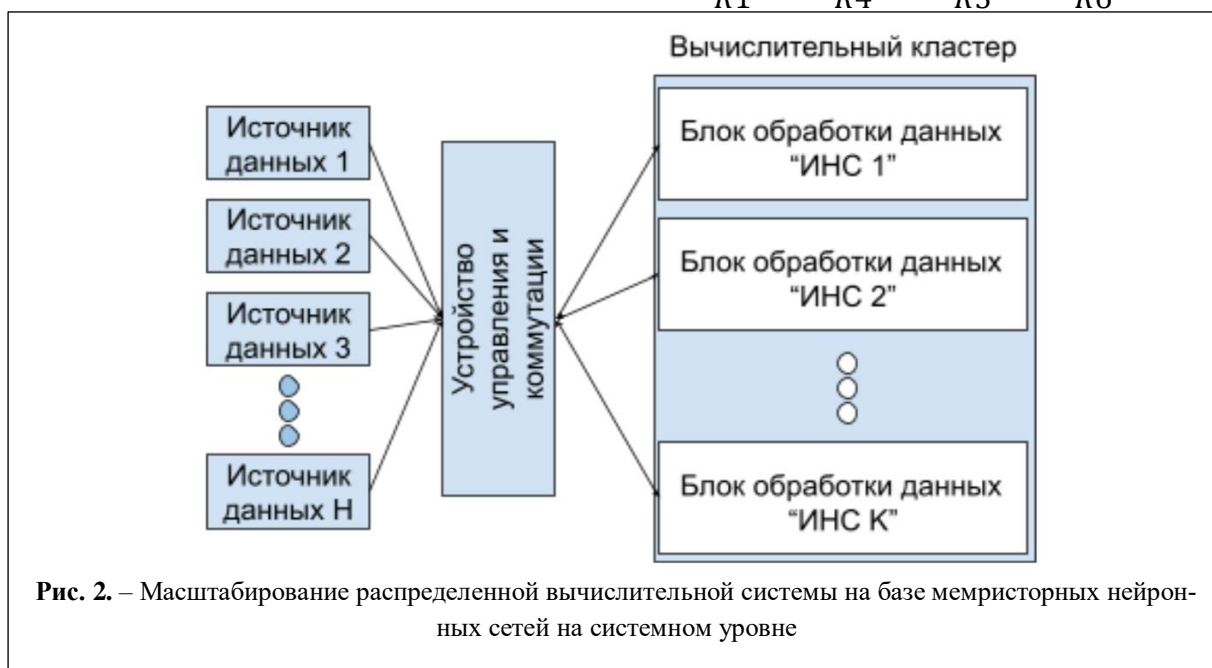
с точки зрения пользователя сможет обслуживать одно или несколько устройств ввода информации, в зависимости от производительности и быстродействия.

В рассматриваемой системе обработка данных происходит в аналоговом виде на кроссбарах мемристивных устройств. Входные данные кодируются амплитудами напряжений входных импульсов. Матрица весов ИНС формируется за счет

изменения сопротивлений мемристивных устройств в кроссбарах (рис. 3)

В соответствии с рис. 3 вес синапса нейрона вычисляется следующим образом:

$$\begin{aligned}
 V_{out} &= V_{oa2} - V_{oa1} = \\
 &= -\frac{R_f}{R_1}V_1 - \frac{R_f}{R_3}V_2 - \frac{R_f}{R_5}V_3 + \\
 &+ \frac{R_f}{R_2}V_1 + \frac{R_f}{R_4}V_2 - \frac{R_f}{R_6}V_3 = \frac{R_f}{R_2}V_1 - \\
 &- \frac{R_f}{R_1}V_1 + \frac{R_f}{R_4}V_2 - \frac{R_f}{R_3}V_2 + \frac{R_f}{R_6}V_3 -
 \end{aligned}$$



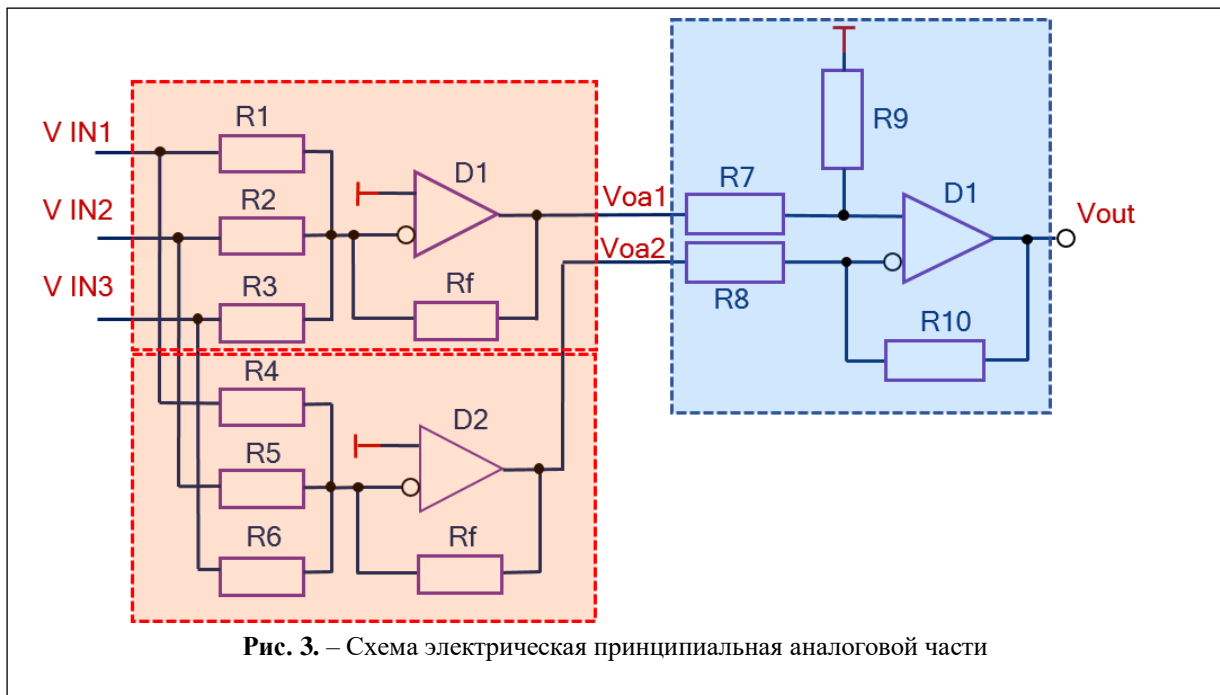


Рис. 3. – Схема электрическая принципиальная аналоговой части

$$-\frac{R_f}{R_5}V_3 = R_f\left(\frac{1}{R_2} - \frac{1}{R_1}\right)V_1 + R_f\left(\frac{1}{R_4} - \frac{1}{R_3}\right)V_2 + R_f\left(\frac{1}{R_6} - \frac{1}{R_5}\right)V_3$$

$$W = R_f \frac{R_2 - R_1}{R_1 R_2},$$

где R_f – сопротивления обратной связи операционных усилителей инвертирующих сумматоров, R – сопротивления мемристивных устройств.

Для моделирования электронных компонентов аналоговой части используется SPICE – язык симулирования электронных схем общего назначения с открытым

исходным кодом, лежащий в основе большинства современных программ для схемотехнического моделирования. Одной из таких программ является LTSpice. Описание электрической принципиальной схемы в LTSpice представляется в виде текстового файла, в котором в каждой строке приводится элемент схемы с указанием его обозначения, номинала и координат (x,y) размещения в схеме (пример на рис. 4).

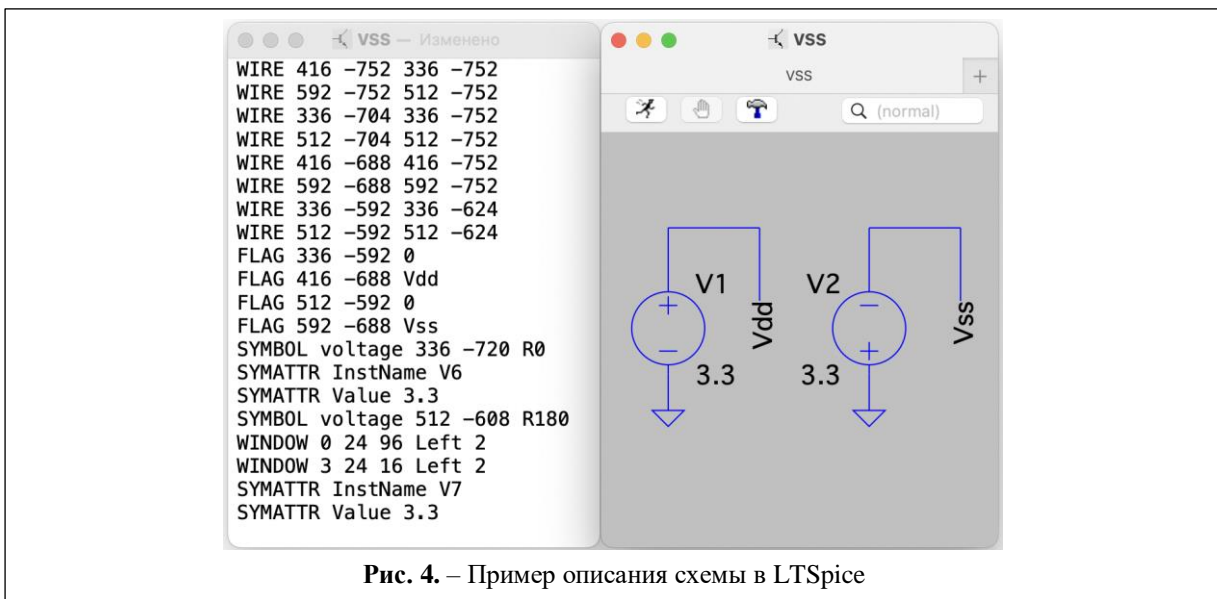
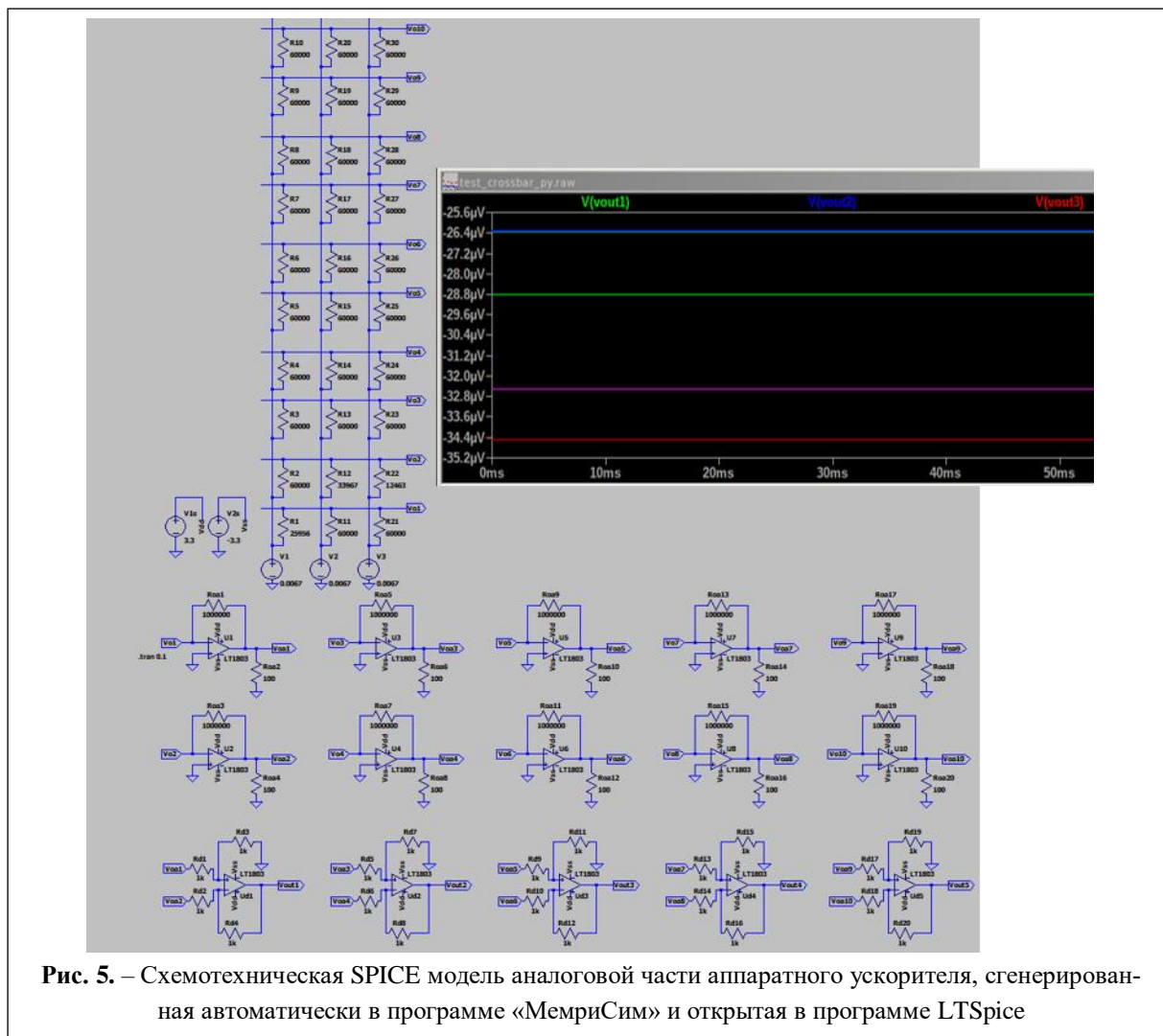


Рис. 4. – Пример описания схемы в LTSpice



2. Моделирование аппаратного ускорителя

Рассматриваемые аппаратные ускорители ИИ имеют в своем составе большое количество параллельно работающих кроссбаров, каждый из которых выполняет МВУ для определенного участка матрицы весов модели ИНС в аналоговом виде. Этот процесс также связан с большим числом операций управления и коммутации, которые не могут быть полноценно реализованы в программах схемотехнического моделирования. В связи с этим авторами предлагается интеграция SPICE моделей, эмулируемых в программе LTSpice в модель системы, эмулирующей работу аппаратного ускорителя модели ИНС в «МемриСим».

Для моделирования логики работы устройства управления и коммутации, управляющего работой всей системы и распределяющего данные между кроссбарами, разработана модель на языке программирования Python. Данная модель, с одной стороны, связана с программной моделью ИНС и источниками входных данных, с другой стороны, эмулируя работу распределенной вычислительной системы путем симулирования автоматически генерируемых моделей LTSpice для аналоговых частей системы. Такая автоматическая генерация осуществляется путем создания текстовых файлов скриптом на Python на основе информации о номинальных значениях весов и входных сигналов, являющихся эквивалентами весов

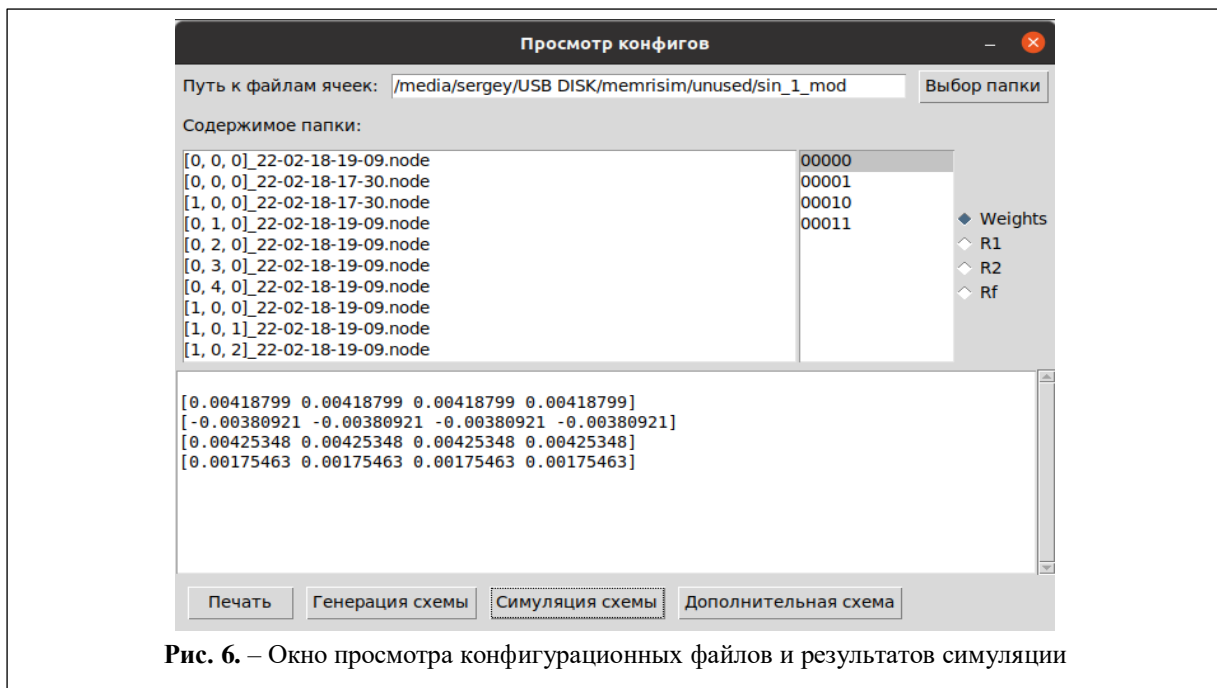


Рис. 6. – Окно просмотра конфигурационных файлов и результатов симуляции

ИНС и входных данных (рис. 5). Симуляция модели LTSpice также запускается из скрипта на Python через команду терминалу операционной системы ЭВМ, на которой выполняется моделирование. Результаты моделирования извлекаются из raw файлов, создаваемых программой LTSpice по окончании симуляции схемы (рис. 6).

Результаты тестирования позволяют сделать вывод о том, что разработанный подход и компьютерная модель позволяют проводить моделирование процесса функционирования распределенной системы обработки информации на базе мемристорных нейронных сетей с учетом того, что это аналогово-цифровая система. Важной особенностью является то, что SPICE модели для аналоговой части генерируются и запускаются полностью автоматически, что значительно снижает временные и трудовые затраты при проведении моделирования.

Заключение

Таким образом в данной работе предложен оригинальный подход к моделированию распределенных систем обработки

информации на базе мемристорных нейронных сетей, основанный на интеграции схемотехнических SPICE моделей, эмулирующих работу аналоговых частей вычислителя в LTSpice на физическом уровне, и моделей на языке программирования Python, эмулирующих работу устройств управления на системном уровне. Предложенный подход программно реализован в виде модуля для ПК «МемриСим».

Литература

1. Mikhaylov A. et al. Multilayer metal-oxide memristive device with stabilized resistive switching // Advanced materials technologies. 2020. Т. 5. №. 1. С. 1900607.
2. Amirsaleimani A. et al. In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal–Oxide–Semiconductor–Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives // Advanced Intelligent Systems. 2020. Т. 2. №. 11. С. 2000115.
3. Галушкин А. И. На пути к нейрокомпьютерам с использованием мемристоров // Информационные технологии. – 2014. – №. 4. – С. 2-19.
4. Борданов И. А., Данилин С. Н., Щаников С. А. Программный комплекс «мемрисим» для имитационного моделирования искусственных нейронных сетей на базе мемристоров // Информационные системы и технологии ИСТ-2021 : сборник материалов XXVII Международной научно-технической конференции Нижегородский государственный технический университет им. Р.Е. Алексеева, Нижний Новгород, 23–24 апреля 2021

года / Нижегородский государственный технический университет им. Р.Е. Алексеева. – Нижний

Новгород: Нижегородский государственный технический университет им. Р.Е. Алексеева, 2021. С. 77-84.

Данная работа выполнялась в рамках решения части задач технического задания на разработку аппаратных ускорителей, финансируемых за счет Субсидии Министерства науки и высшего образования РФ (проект № 13.2251.21.0098, соглашение № 075-15- 2021-1017).

Поступила 14 ноября 2021 г.

The article is devoted to solving an important scientific and practical problem related to the development of methods and hardware and software tools for modeling distributed information processing systems based on memristive neural networks. The paper presents an original approach to modeling such systems based on the integration of circuit SPICE models that emulate the operation of analog parts of the computer in LTspice at the physical level, and models in the Python programming language that emulate the operation of control devices at the system level.

Key words: artificial neural networks, memristors, distributed data processing, hardware accelerators of artificial intelligence.

Никишов Даниил Андреевич – инженер кафедры информационных систем Муромского института (филиала) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: daniilnikisov74@gmail.com.

Антонов Александр Михайлович – инженер кафедры информационных систем Муромского института (филиала) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: cfifant@mail.ru.

Жиганова Елена Александровна – кандидат технических наук, доцент кафедры радиотехники Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: zhiganova.el@gmail.com.

Жиганов Сергей Николаевич – кандидат технических наук, доцент кафедры радиотехники Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: s_zh_72@mail.ru.

Борданов Илья Алексеевич – инженер кафедры информационных систем Муромского института (филиала) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: bordanov2011@yandex.ru.

Данилин Сергей Николаевич – кандидат технических наук, доцент кафедры программно-инженерии Муромского института (филиала) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: dsn-55@mail.ru.

Щаников Сергей Андреевич – кандидат технических наук, декан факультета информационных технологий Муромского института (филиала) федерального государственного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых».

E-mail: seach@inbox.ru.

Адрес: 602264, Муром, ул. Орловская, д. 23.