

На пути к нейрокомпьютерам с использованием мемристоров

(рабочий вариант)

Оглавление

1. Введение.....	1
2. Нейронные сети.....	3
3. Нейрокомпьютер.....	9
4. Нейрочипы.....	12
5. Эффективные программные эмуляторы нейросетевых алгоритмов.....	14
6. Аппаратная реализация нейросетевых алгоритмов решения задач.....	17
7. Перспективы развития нейрокомпьютеров в XXI веке.....	19
8. Аналоговая обработка – принципиальное свойство мемристорных систем.....	20
9. Особенности архитектуры современных и перспективных суперЭВМ.....	21
10. Научно-технический задел по созданию отечественных супернейрокомпьютеров.....	23
11. Мемристоры – детище нанотехнологий.....	26
12. Первоочередные работы в области мемристорных систем.....	29
13. Анализ технологий изготовления мемристоров и мемристорных систем.....	29
14. Замечание к оценке производительности.....	30
15. Сфера применения будущих мемристорных систем.....	32
16. Архитектура супернейрокомпьютера с использованием мемристоров.....	32
17. Базовые технические средства.....	33
18. Системное программное обеспечение.....	34
19. Оценка качества вычислительных систем с использованием мемристоров.....	35
20. Состав работ ОКР по созданию образцов супернейрокомпьютера на базе мемристоров (2016-2018 гг.).....	35
21. Разработка прикладного программного обеспечения супернейрокомпьютера с использованием мемристоров.....	35
22. Нейропротезы – эффективное направление применения НБИКС технологий.....	38
23. О роли нейрофизиологических исследований.....	38
24. НБИКС технологии и проблема искусственного интеллекта.....	40
Заключение.....	43
Литература.....	45

1. Введение.

Принципиально новые возможности, возникающие на стыке нанотехнологий, биотехнологий, информационных технологий, когнитивных технологий и социо-гуманитарных технологий (НБИКС), обусловили возникновение в XXI веке новой конвергентной научной и технологической парадигмы – НБИКС-технологии. Основными чертами НБИКС-технологии является переход к биоподобным энергосберегающим решениям за счет конвергенции передовых достижений в различных смежных дисциплинах.

Нейросетевые технологии были рождены в середине 40ых годов прошлого столетия работами известного нейрофизиолога Маккалока У.С. в процессе проведения под его председательством 10-и конференций (Cerebral Inhibition Meeting), в частности его совместной с Уолтером Питтсом работой «Логическое исчисление идей, относящихся к нервной активности» (1943) [1]. Эти технологии в последующие десятилетия получили активное развитие в работах ряда зарубежных (Ф. Розенблатт, Р. Хехт-Нильссен, Лотфи Заде, П. Вербос и др.) и отечественных ученых (В.М. Глушков, А.Г. Ивахненко, В.В. Солодовников, Я.З. Цыпкин и др.)

Основой БИК технологий всегда являлись нейросетевые технологии, которые в рамках создания конкретных интеллектуальных информационных систем в 50ые, 60ые и 70ые годы прошлого столетия применялись в основном для решения различных задач распознавания образов. В связи с этим получила широкое развитие теория нейронных сетей, включающая в себя разработку структур нейронных сетей и различных алгоритмов настройки коэффициентов связей нейронов в этих сетях

Нейросетевые технологии в рамках создания конкретных интеллектуальных информационных систем в 50ые, 60ые и 70ые годы прошлого столетия применялись в основном для решения различных задач распознавания образов. В связи с этим получила широкое развитие теория нейронных сетей, включающая в себя разработку структур нейронных сетей и различных алгоритмов настройки коэффициентов связей нейронов в этих сетях.

Направление создания искусственных нейросетей зародилось в двадцатом веке на границе математики и нейрофизиологии благодаря технологиям электроники. Вплоть до начала двадцать первого века нейрокомпьютеры развивались главным образом благодаря достижениям микроэлектроники. В тоже время разрабатывался математический аппарат – нейроматематика.

В двадцать первом веке тенденции существенно изменились. Возросшие мощности суперкомпьютеров дают возможность ставить задачу симуляции нейросистем живых существ, что фактически означает возвращение в направление искусственных нейронных сетей когнитивной науки. Развитие этих работ должно привести не только к более глубокому пониманию перехода от микро- к

макрокогнитивной функции живых существ, но и большому числу прикладных применений.

Другой тенденцией стали активные разработки новых аппаратных реализаций аналоговых нейрочипов с частотной модуляцией на основе гибридизации КМОП-технологии и мемристоров [2], ставшие возможными при переходе к нанотехнологиям. Такая парадигма реализации искусственных нейросетей гораздо ближе к механизмам функционирования живых нейронов, чем цифровые или аналоговые на постоянном токе. В этом направлении проявляются все признаки парадигмы НБИКС-технологий: повышение производительности, понижение энергопотребления и уменьшение размера устройства за счет копирования свойств живой природы.

Следует подчеркнуть, что на пороге второго десятилетия двадцать первого века резко возрос интерес крупных индустриальных держав (США, ЕС), а также высокотехнологических компаний (Intel, IBM, Hewlett Packard и др.) к искусственным нейросетям и разработке новых нейрокомпьютеров. На кону очередного этапа технологической гонки широкий спектр возможностей от прикладных разработок в гражданских и военных областях до фундаментальных знаний о механизмах когнитивной функции животных и человека и создания систем искусственного интеллекта.

Однако, несмотря на возросший в последние годы интерес к искусственным нейронным сетям, история развития этого направления имела скорее поступательный, нежели революционный характер. Поэтому понимание основных шагов программной и аппаратной реализации нейронных сетей является ключом к созданию нейросетевых решений нового поколения [3]. Ниже определяются основные понятия, связанные с искусственными нейронными сетями, определяются задачи, эффективно решаемые нейрокомпьютерами, приводятся исторические этапы нейросетевых программных и аппаратных реализаций. В конце представлены возможные пути развития технологии нейрочипов, в том числе на базе мемристоров, а также пути реализации систем на их основе, делая основной упор на реализацию супернейрокомпьютеров.

2. Нейронные сети

Нейронная сеть - искусственная, многослойная высокопараллельная (т.е. с большим числом независимо параллельно работающих элементов) логическая структура, составленная из формальных нейронов. Появление формального нейрона во многом обусловлено изучением биологических нейронов. Формальный нейрон представляет собой логический элемент с N входами, $(N+1)$ весовыми коэффициентами, сумматором и нелинейным преобразователем. Простейший формальный нейрон, осуществляющий логическое преобразование

$$y = \text{sign} \sum_{\substack{i=0 \\ x_0=1}}^N a_i x_i,$$

входных сигналов (которыми например, являются выходные сигналы других формальных нейронов нейронной сети) в выходной представлен на рис. 1, где y – значение выхода формального нейрона; a^i – весовые коэффициенты; x^i – входные значения формального нейрона ($x^i \in \{0,1\}$).

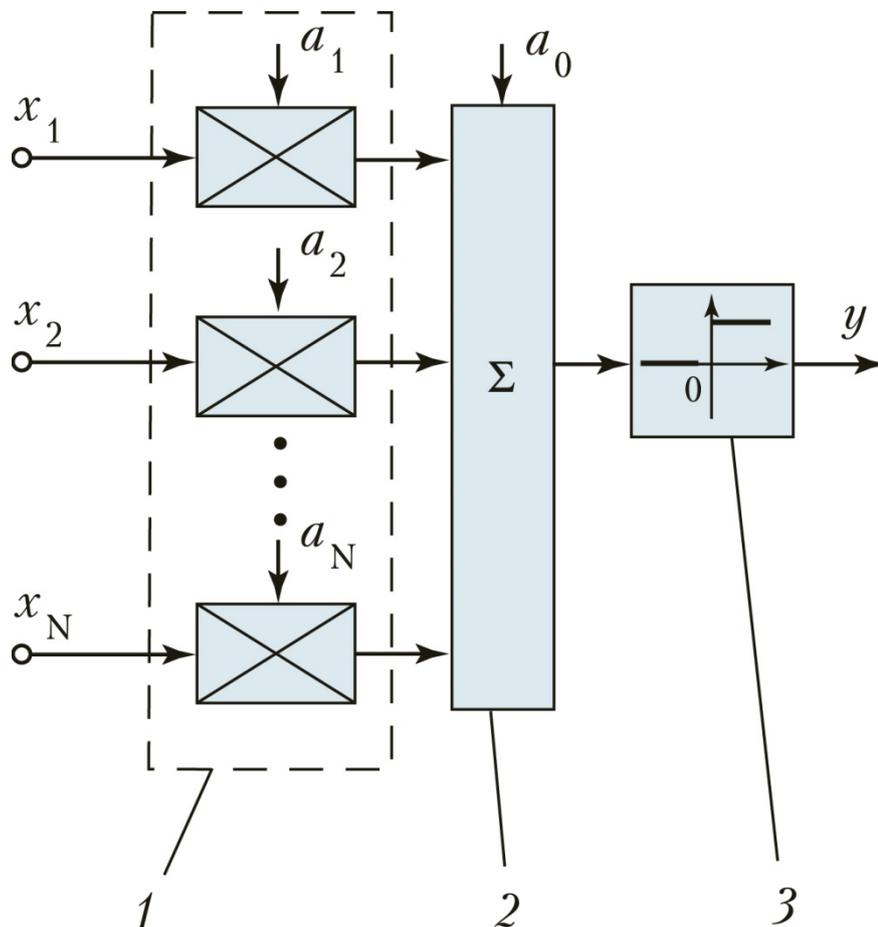


Рис.1. Функциональная схема нейрона. 1 – блок входа; 2 – сумматор; 3 – блок нелинейного преобразования.

Процесс вычисления выходного значения формального нейрона представляет собой движение потока данных и их преобразование. Сначала данные поступают на блок входа формального нейрона, где происходит умножение исходных данных на соотв. весовые коэффициенты. Весовой коэффициент является мерой, которая определяет насколько соответствующее входное значение влияет на состояние формального нейрона. Весовые коэффициенты могут изменяться в соответствии с обучающими примерами, архитектурой нейронной сети, правилами обучения и др. Полученные (при умножении) значения преобразуются в сумматоре в одно числовое значение g (посредством суммирования). Затем, для определения выхода формального нейрона в блоке нелинейного преобразования (реализующего передаточную функцию) g сравнивается с некоторым числом (порогом). Если сумма больше значения порога, формальный нейрон генерирует сигнал, в противном случае сигнал будет нулевым или тормозящим. В данном формальном нейроне применяется нелинейное преобразование:

$$\text{sign}(g) = \begin{cases} 0, & g < 0 \\ 1, & g \geq 0 \end{cases}$$

На рис.2 показана общая схема многослойной нейронной сети, с последовательными связями. Высокий параллелизм обработки достигается путём объединения большого числа формальных нейронов в слои и соединения определённым образом различных нейронов между собой.

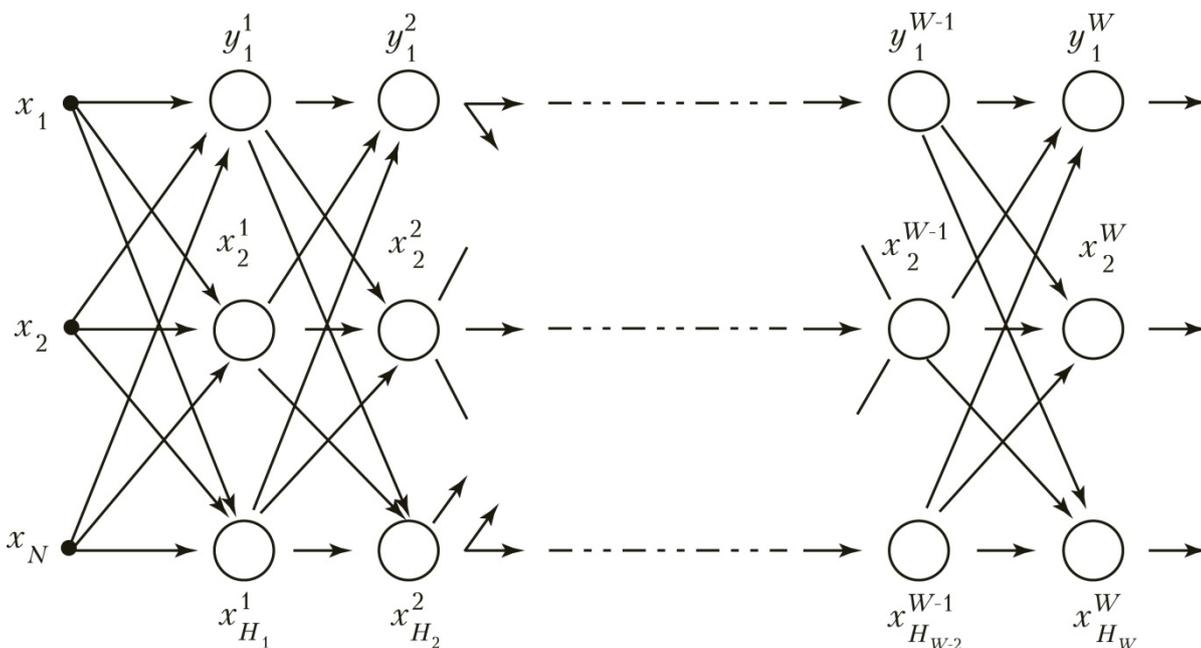


Рис. 2. Граф многослойной нейронной сети с последовательными связями. x^i – i -й входной сигнал ($i = 1..N$); H^1 – число нейронов 1-го слоя; W – число слоёв; x^{H^W} – выходной сигнал H^W -го нейрона W -го слоя; y^1 – выходное значение W -го слоя.

В общем случае в эту структуру могут быть введены перекрёстные и обратные связи с настраиваемыми весовыми коэффициентами (рис.3).

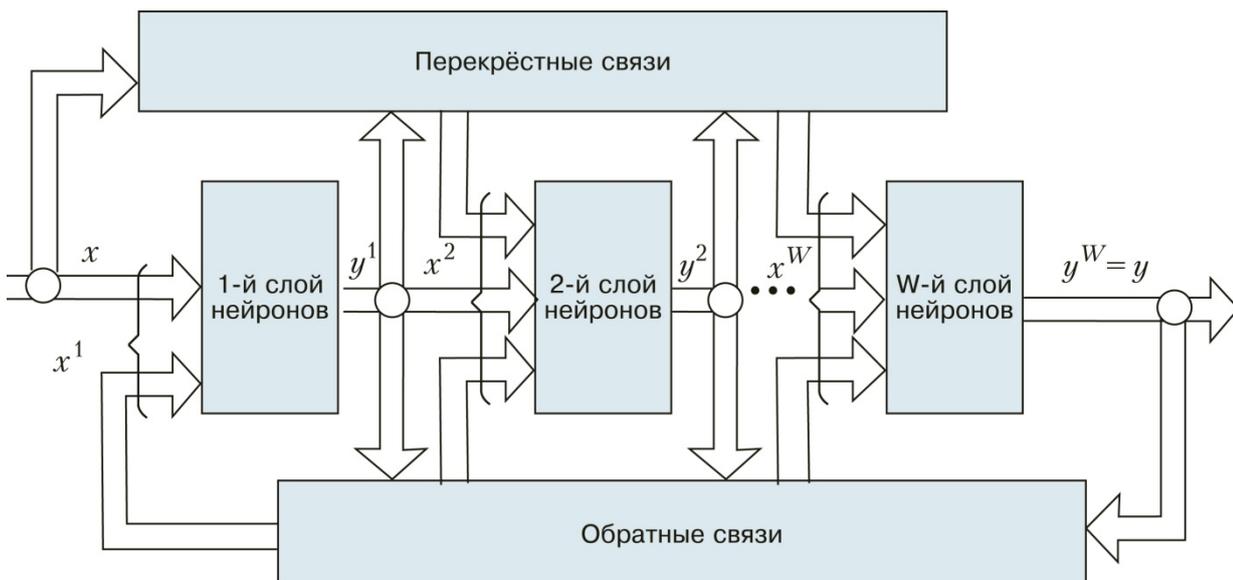


Рис.3. Многослойная нейронная сеть.

На дальнейшее развитие теории нейронных сетей существенное влияние оказала монография американского нейрофизиолога Ф. Розенблатта «Принципы нейродинамики» (1962) [3], в которой он предложил схему *перцептрона* (устройства, моделирующего процесс восприятия информации человеческим мозгом). Его идеи получили развитие в большом числе научных работ других авторов. Теория нейронных сетей продолжает достаточно активно развиваться и в начале 21 века [4, 5].

Нейронные сети являются сложными нелинейными системами с огромным числом степеней свободы. Принцип, по которому они обрабатывают информацию, отличается от принципа, используемого в компьютерах на основе процессоров с фон-неймановской архитектурой (с логическим базисом И, ИЛИ, НЕ). Вместо классического программирования (как в традиционных вычислительных системах) применяется обучение нейронных сетей, которое сводится, как правило, к настройке весовых коэффициентов с целью оптимизации заданного критерия качества функционирования нейронной сети. Нейросетевым алгоритмом решения задач называется вычислительная процедура, основная часть которой или она целиком реализована в виде нейронной сети той или иной структуры (например, многослойная нейронная сеть с последовательными или перекрёстными связями между слоями формальных нейронов) с соответствующим алгоритмом настройки весовых коэффициентов. Основой разработки нейросетевого алгоритма является системный подход, при котором процесс решения задачи представляется как функционирование во времени некоторой динамической системы. Для её построения необходимо определить: объект, выступающий в роли входного сигнала нейронной сети; объект, выступающий в роли выходного сигнала нейронной сети (например, непосредственно решение или некоторая его характеристика); желаемый (требуемый) выходной сигнал нейронной сети; структуру нейронной сети (число слоёв, связи между слоями, объекты, служащие весовыми коэффициентами); функцию ошибки системы (характеризующую отклонение желаемого выходного сигнала нейронной сети от реального выходного сигнала); критерий качества системы и функционал её оптимизации, зависящий от ошибки; значение весовых коэффициентов (например, определяемых аналитически

непосредственно из постановки задачи, с помощью некоторых численных методов или процедуры настройки весовых коэффициентов нейронной сети).

Количество и тип формальных нейронов в слоях, а также число слоёв нейронов выбираются исходя из специфики решаемых задач и требуемого качества решения. Нейронная сеть в процессе настройки на решение конкретной задачи рассматривается как многомерная нелинейная система, которая в итерационном режиме целенаправленно ищет оптимум некоторого функционала, количественно определяющего качество решения поставленной задачи. Для нейронных сетей, как многомерных нелинейных объектов управления, формируются алгоритмы настройки множества весовых коэффициентов. Основные этапы исследования нейронных сетей и построения алгоритмов настройки (адаптации) их весовых коэффициентов включают: исследование характеристик входного сигнала для различных режимов работы нейронной сети (входным сигналом нейронной сети является, как правило, входная обрабатываемая информация и указание т.н. учителя нейронной сети); выбор критериев оптимизации (при вероятностной модели внешнего мира такими критериями могут быть минимум средней функции риска, максимум апостериорной вероятности, в частности при наличии ограничений на отдельные составляющие средней функции риска); разработка алгоритма поиска экстремумов функционалов оптимизации (например, для реализации алгоритмов поиска локальных и глобального экстремумов); построение алгоритмов адаптации коэффициентов нейронной сети; анализ надёжности и методов диагностики нейронной сети и др.

Достоинствами нейронных сетей являются: свойство т.н. постепенной деградации – при выходе из строя отдельных элементов качество работы системы падает постепенно (для сравнения, логические сети из элементов И, ИЛИ, НЕ выходят из строя при нарушении работы любого элемента сети); повышенная устойчивость к изменению параметров схем, их реализующих (например, весьма значительные изменения весовых коэффициентов не приводят к ошибкам в реализации простой логической функции 2-х переменных) и др.

Широкое распространение нейросетевых алгоритмов в области сложных формализуемых, слабоформализуемых и неформализуемых задач привело к созданию нового направления в вычислительной математике – нейроматематики

[6]. Нейроматематика включает нейросетевые алгоритмы решения следующих задач: распознавание образов; оптимизация и экстраполяция функций; теория графов; криптографические задачи; решение вещественных и булевских систем линейных и нелинейных уравнений, обыкновенных одномерных и многомерных дифференциальных уравнений, дифференциальных уравнений в частных производных и др.

На основе теории нейронных сетей создан новый раздел современной теории управления сложными нелинейными и многомерными, многосвязными динамическими системами, включающий, методы нейросетевой идентификации сложных динамических объектов; построение нейрорегуляторов в контурах управления сложными динамическими объектами и др [7].

3. Нейрокомпьютер

Нейрокомпьютер (от гр. *neuron* – нерв и *компьютер*), вычислительная система, в которой аппаратное и программное обеспечение оптимизировано для реализации алгоритмов решения задач на основе принципов работы нейронных сетей [8]. Появление нейрокомпьютеров обусловлено прежде всего отказом от логического базиса И, ИЛИ, НЕ, используемого в компьютерах на основе процессоров с фон-неймановской архитектурой, сначала на уровне описания алгоритмов решения задач, затем на уровне элементной базы вычислительных машин с постепенным переходом на использование в качестве базовых элементов вычислительной системы формальных нейронов (с соответствии алгоритмами настройки весовых коэффициентов входных сигналов), которые в простейшем случае моделируют функции биологических нейронов. Нейросетевые алгоритмы решения задач обеспечивают максимально возможный уровень распараллеливания вычислений при современной аппаратной реализации по сравнению с др. алгоритмами. Нейрокомпьютеры применяют в основном для решения общесистемных задач (обработка сигналов – речевых, сейсмических, радиолокационных, гидроакустических и др.; обработка изображений – биометрические, аэрокосмические и др.; реализация систем принятия решений, систем защиты информации и др.); задач промышленного назначения (в геоинформационных, телекоммуникационных, космических, навигационных

системах, для диагностики технических систем и др.), социально-экономических задач (в медицине, биоинформатике, метеорологии, образовании, экономике и др.). Основные преимущества нейрокомпьютеров: параллелизм вычислений, что обуславливает высокое быстродействие; низкие требования к стабильности и точности параметров элементарных узлов; устойчивость к помехам и разрушениям (выходу из строя) формальных нейронов при большой пространственной размерности системы (причём устойчивые и надёжные нейронные сети могут создаваться из малонадёжных элементов, имеющих большой разброс параметров).

Во всем мире разрабатывались десятки типов нейрокомпьютеров, специализированных вычислительных систем с нейросетевой архитектурой, принципиально отличной от архитектуры фон-Неймана, принятой в то время в универсальных цифровых вычислительных машинах.

Аппаратные разработки нейрокомпьютеров начались в 1950-е гг. и велись параллельно с работами над универсальными ЭВМ; постепенно сфера применения нейрокомпьютеров расширялась вследствие роста числа сложных формализуемых и неформализуемых задач, реализовать решение которых на классических компьютерах невозможно или очень трудно. Первый в мире нейрокомпьютер «MARK 1» или *перцептрон* создан в 1958 американским нейрофизиологом Ф. Розенблаттом (рис. 4). Ряд специализированных и экспериментальных нейрокомпьютеров разрабатывался в 1960 – 70-е гг. 20 века (рис. 5). В конце 1980-х гг. активизировались разработки цифровых нейрокомпьютеров, что связано с развитием микроэлектроники, напр., стало возможно реализовать в одном кристалле фрагмент нейронной сети с каскадным соединением формальных нейронов.

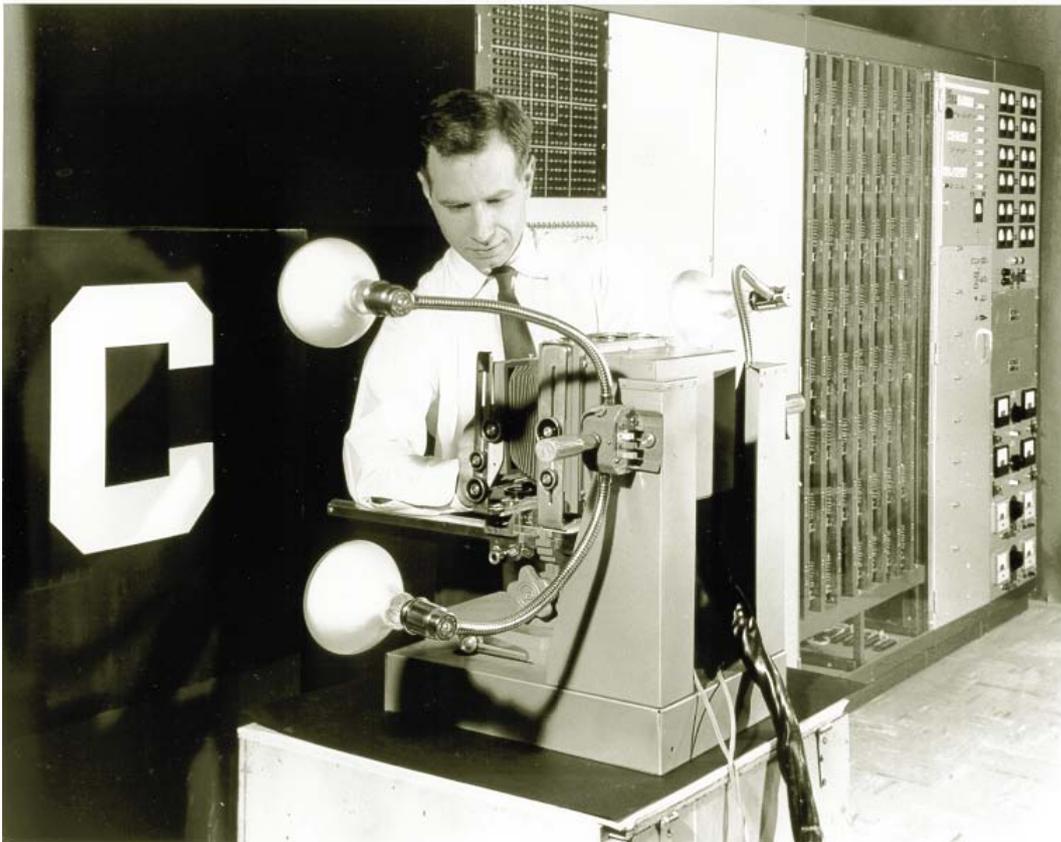


Рис.4. Персептрон Марк-1 в Корнельской лаборатории аэронавтики. 1958г.



Рис. 5. Общий вид отечественного нейрокомпьютера 1974 г. Размерность входного пространства признаков 30. Количество распознаваемых классов 8.

По способам обработки сигналов различают цифровые, аналоговые и аналого-цифровые нейрокомпьютеры (именно нейрокомпьютеры обусловили тенденцию возврата к аналоговым методам обработки информации, обладающим высоким быстродействием и низкой стоимостью). Известны реализации нейрокомпьютеров в виде: совокупности *нейрочипов* (цифровых, аналоговых, аналого-цифровых); встраиваемых блоков, аппаратно реализующих различные физические принципы действия (оптические, молекулярные, квантовые) [9,10,11]. К нейрокомпьютерам также относятся рабочие станции или суперЭВМ (на основе процессоров с фон-неймановской архитектурой, т.е. с логическим базисом И, ИЛИ, НЕ) различной архитектуры с программным обеспечением, реализующим нейросетевые алгоритмы решения задач. В начале 21 века наиболее востребованы цифровые нейрокомпьютеры, которые реализуют в различных технологических вариантах, например, на базе ПЛИС – программируемых логических интегральных схем (логических матриц), т.н. заказных цифровых нейрочипов, систем на кремниевой пластине.

Разработки теории нейронных сетей, нейроматематики и нейрокомпьютеров тесно взаимосвязаны. С одной стороны развитие нейроматематики способствует развитию теории нейронных сетей, которая в свою очередь направляет развитие технических разработок нейрокомпьютеров, с другой – уровень развития технических средств определяет диапазон возможностей для построения нейронных сетей и нейросетевых алгоритмов решения задач.

4. Нейрочипы

Нейрочип - интегральная микросхема, в архитектуре, системе команд и программном обеспечении которой заложены некоторые характеристики нейронных сетей. Нейрочип аппаратно представляет связанное множество элементов умножения сигналов, их суммирования и нелинейного преобразования (т.е. формальный нейрон). Как правило, нейрочип реализует небольшой фрагмент *нейронной сети* и обладает свойством т.н. каскадирования, т.е. образования

нейронной сети большего размера с размещением на одной печатной плате нескольких нейрочипов под единым управлением. Известно несколько сотен типов нейрочипов с размещением на одном кристалле от нескольких штук до нескольких миллионов формальных нейронов. С развитием технологий микроэлектроники интеграция нейрочипов постоянно растёт. Начало разработок нейрочипов относится к концу 1960-х – началу 70-х гг., когда появились первые цифровые интегральные схемы средней интеграции. Активное развитие нейрочипы получили после создания больших и сверх больших интегральных схем (СБИС), особенно, когда технология СБИС позволила на одном кристалле реализовать фрагмент нейронной сети значительного размера. По технологии реализации большинство нейрочипов подразделяют на цифровые [12,13] и аналоговые (аналого-цифровые) [14,15,16,17]. По назначению нейрочипы делят на универсальные (предназначенные для решения широкого класса задач) и специализированные (ориентированные на решение определённого класса задач). Многие попытки сделать универсальные нейрочипы в 1988 – 2010 были малоэффективными, т.к. нейроматематика на данном этапе ещё недостаточно развита. Наиболее широкое распространение получили специализированные нейрочипы, которые обеспечивали максимальную эффективность решения конкретной задачи, например, обработка сигналов, изображений (сжатие, выделение движущихся объектов и др.), управление роботами, реализация ассоциативной памяти, осуществление аналого-цифрового преобразования. При этом и в цифровом и в аналого-цифровом вариантах в нейрочипе реализуются фрагменты нейронных сетей, например, многослойных, рекуррентных (с обратными связями между элементами) и др. Особо выделяют нейрочипы, в которых входные, выходные и промежуточные сигналы представляют собой последовательность достаточно узких импульсов с переменной частотой следования, которая и является основной информационной составляющей.

Большинство цифровых нейрочипов (т.н. полузаказные нейрочипы) последние 20 лет реализовывались на базе программируемых логических интегральных схем (ПЛИС) [18]. Разрабатывались и применялись т.н. заказные цифровые нейрочипы, в основном в США, Японии, Франции.

Для технологий реализации нейрочипов и *нейрокомпьютеров* характерен активный переход от цифровых технологий к аналоговым и аналого-цифровым. Отметим, что в 1960 – 70-е гг. переход от аналоговых технологий к цифровым произошёл в основном из-за недостаточной точности вычислений. В аналоговых и аналого-цифровых нейрочипах и нейрокомпьютерах этот недостаток компенсируется высокой параллельностью алгоритмов обработки данных и специализацией вычислительных систем на их базе.

Уровень развития технологий микроэлектроники, значительный рост степени интеграции элементной базы, естественная необходимость перехода к более однородным схемам реализации, нежели классические фон-неймановские процессоры, обусловили в начале 21 века новый этап развития цифровых и аналоговых технологий нейрочипов. К перспективным технологиям в первую очередь относятся технологии реализации мемристорных, квантовых и молекулярных нейрочипов и нейрокомпьютеров.

5. Эффективные программные эмуляторы нейросетевых алгоритмов

Изучение формальных моделей нейронных сетей уже более 70 лет занимает умы человечества. И всегда для изучения поведения нейросетевых алгоритмов использовались самые совершенные из доступных вычислительных систем. До середины 80ых годов прошлого столетия это были одно- и микропроцессорные ЭВМ и миниЭВМ. С появлением в начале 80ых годов прошлого столетия транспьютерных систем [19] наступила эра применения для программной эмуляции нейросетевых алгоритмов так называемых масштабируемых ЭВМ, в которых производительность может наращиваться с ростом числа процессоров.

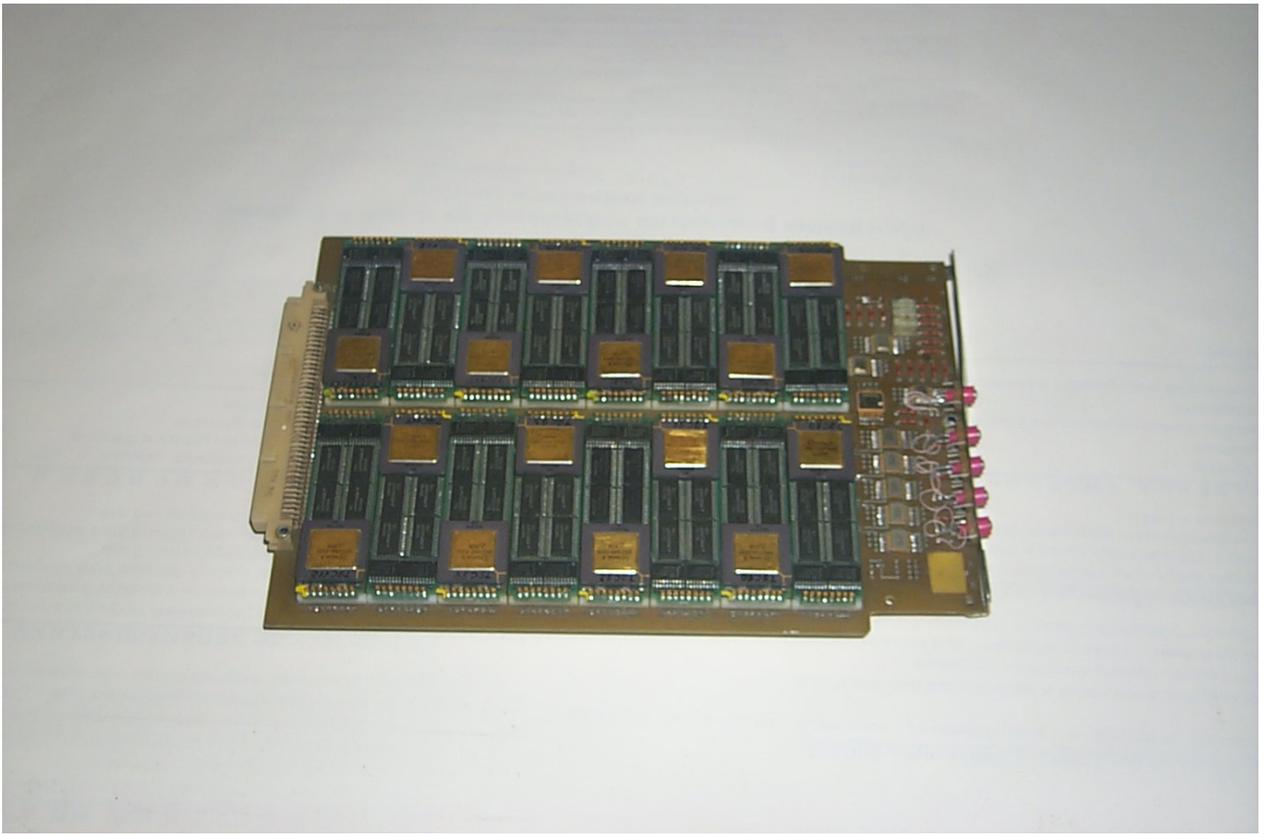


Рис. 6. Общий вид отечественной транспьютерной платы (32 транспьютерных модуля)

Самой совершенной из транспьютерных систем была система на базе транспьютеров Т9000 (рис. 7).

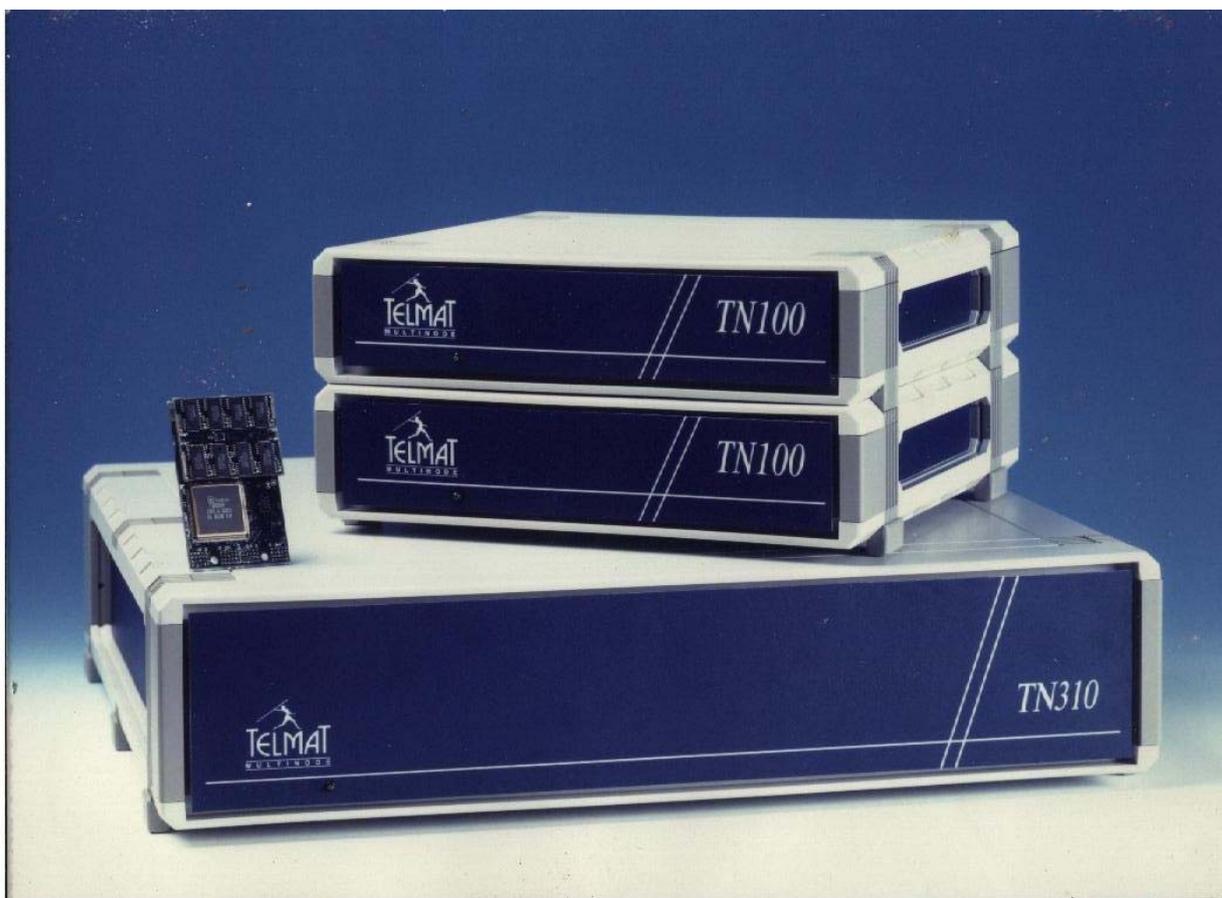


Рис. 7. Общий вид транспьютерной системы на базе транспьютера T9000

Чисто транспьютерные системы, состоящие только из транспьютеров, оказались малоэффективными для решения многих задач, в том числе для эмуляции нейросетевых алгоритмов. Именно по этой причине были разработаны транспьютерные сопроцессоры обработки сигналов (IMS A100), изображений (IMS A110), представляющие собой каскадируемые Z-фильтры с перестраиваемыми в процессе решения задач коэффициентами. С их применением транспьютерные системы реализовывали двухслойные масштабируемые архитектуры, которые явились тогда основой будущих и современных суперЭВМ (в том числе на базе ПЛИС и GPU) и являются необходимым свойством будущих суперЭВМ, в том числе с применением мемристоров.

В настоящее время самыми эффективными эмуляторами нейросетевых алгоритмов являются системы на ПЛИС (FPGA) и суперЭВМ на базе графических процессоров (GPU).

6. Аппаратная реализация нейросетевых алгоритмов решения задач

Подобно классической универсальной вычислительной технике, нейрокомпьютеры реализовывались на базе цифровых и аналоговых технологий. В 50ые и 60ые года прошлого столетия низкая производительность цифровых вычислительных машин не позволяла реализовывать достаточно быстродействующие нейрокомпьютеры с большим числом нейронов и настраиваемых коэффициентов.

Именно поэтому нейрокомпьютеры реализовывались с применением аналоговых технологий в двух направлениях:

- с аналоговой аппаратной реализацией контура настройки весовых коэффициентов с применением адаптивных элементов аналоговой памяти (в частности электрохимических элементов – мемисторов);

- с аналоговой реализацией нейронной сети с ручной настройкой коэффициентов и программной реализацией на цифровой машине контура настройки коэффициентов.

Первое направление аналоговой реализации не было развито из-за малой в то время скорости переключения связей (весовых коэффициентов нейронной сети).

Второе направление нашло свои применения для задач со стабильным некоторое время архивом исходных данных для обучения и возможностью регулярной перестройки коэффициентов через некоторые промежутки времени, в течение которых архив данных для обучения пополнялся.

Второе направление прекратило свое существование в 70е годы прошлого столетия с появлением микропроцессоров, миниЭВМ и высокопроизводительных больших ЭВМ.

С начала 90ых годов прошлого столетия реализовано несколько десятков типов нейрокомпьютеров с применением цифровых и аналоговых технологий СБИС. Основными разработками нейрокомпьютеров на базе цифровых технологий были:

- нейрокомпьютер CNAPS;
- нейрокомпьютеры фирмы Sundance;
- нейрокомпьютеры на базе нейрочипов L-neuro (фирма Philips);

- нейрокомпьютеры на базе нейрочипов ZISC (zero instruction set computer – компьютер с нулевым набором команд) фирмы IBM.

Эти нейрокомпьютеры реализовывались на базе заказных цифровых нейрочипов.

Достаточно большое количество цифровых нейрокомпьютеров разрабатывалось и разрабатывается в настоящее время на базе цифровых программируемых логических интегральных схем.

Кардинальным моментом в развитии нейрокомпьютеров является возврат к аналоговой реализации, произошедшей вследствие развития нанотехнологий в микроэлектронике.

Этот возврат характерен для вычислительных систем с нейросетевой архитектурой алгоритмов и аппаратного обеспечения и ставит несколько целей:

- резкое повышение производительности вычислений за счет использования аналоговой обработки во втором слое (оболочке) вычислительной системы.

- понижение энергопотребления за счет перехода в представлении информации в виде уровней токов и напряжений к представлению информации в виде частоты последовательности узких электрических импульсов.

Примечание. Эта попытка была сделана в 70ые-80ые годы прошлого столетия [20] для вычислительных систем с фон-неймановской архитектурой и привела к достаточно эффективному снижению энергопотребления и повышению надежности, но в целом не привела к успеху.

- упрощение процесса проектирования СБИС за счет перехода от неоднородных схем, характерных для фон-неймановских архитектур, к однородным схемам, характерным для нейронных сетей.

За последние 20 лет разработано несколько десятков типов аналоговых нейрочипов по технологии 3.0-0.8 мкм. с классической схемотехникой, характерной для аналоговых и аналогово-цифровых микросхем. Наиболее характерным для этого периода времени является нейрокомпьютер фирмы Intel на базе аналоговых нейрочипов ETANN (Electrically Trainable Artificial Neural Network). Разработки в области нейрокомпьютеров фирма Intel продолжает до сих пор, в том числе на базе мемристоров.

7. Перспективы развития нейрокомпьютеров в XXI веке.

Дальнейшее развитие аппаратной реализации нейрочипов будет преследовать целый ряд задач:

- Увеличения количества эмулируемых нейронов;
- Увеличение количества связей между нейронами;
- Увеличение быстродействия системы;
- Устойчивость работы системы;
- Возможность каскадирования нейрочипов для создания суперкомпьютеров.

Огромные перспективы для реализации перечисленных задач были открыты с созданием в 2008 году сотрудниками фирмы Hewlett Packard мемристоров на основе оксида титана [21]. Мемристор – это элемент, изменяющий свое сопротивление под действием тока, прошедшего через него. Таким образом, этот элемент с эффектом памяти может выступать в роли синапса или весового коэффициента нейронной сети [22, 23]. Являясь близким родственником мемристоров, мемристор интересен для использования в современных нейрочипах благодаря компактности реализации (работают элементы размером десятки нанометров) и скорости переключения, теоретически достигающей до наносекунд [24].

Фактически появление мемристоров дает возможность разработки аналоговых нейрочипов с высоким быстродействием. На настоящий момент нет единой концепции создания нейрочипов с использованием мемристоров. Hewlett Packard пошло по пути совмещения КМОП-технологии с мемристорными матрицами [25]. Intel использует совмещение матрицы мемристоров со спиновыми вентилями [26]. Существующее на настоящий момент многообразие подходов к конструированию нейрочипов нового поколения говорит о широком потенциале новых мемристорных элементов.

Еще одной принципиальной особенностью новых разработок нейрочипов с мемристорами должна стать частотная модуляция сигнала. Такой подход не только позволит существенно снизить энергопотребление чипов, но и существенно повысить срок службы и надежность элементов.

Таким образом, аппаратные реализации нейрочипов нового поколения могут быть получены на основе мемристоров. Отличительными чертами таких нейрочипов станут компактность, высокое быстродействие и низкое энергопотребление [27].

8. Аналоговая обработка – принципиальное свойство мемристорных систем

На рис. 8 условно представлена динамика развития аналоговых и цифровых технологий в части вычислительных систем. Отмечено активное развитие АВМ – аналоговых вычислительных машин в период 50ых-60ых годов прошлого столетия, когда цифровые вычислительные машины (ЦВМ) имели достаточно низкую производительность.

Первые нейрокомпьютеры в 60ых и начале 70s[годов прошлого века создавались на базе аналоговых ЭВМ МН7, МН14 и других.

Активный рост производительности ЦВМ, а также появление микропроцессоров в 70е годы прошлого столетия привел к затуханию разработок и применения АВМ.

В настоящее время известные ограничения цифровых технологий и развитие нанотехнологий делает возможным возврат к аналоговой обработке на фоне высокопараллельных нейросетевых алгоритмов и вычислительных структур с определением адекватного места применения аналоговых технологий в структуре вычислительной системы и соответствующим контролем точности вычислений.

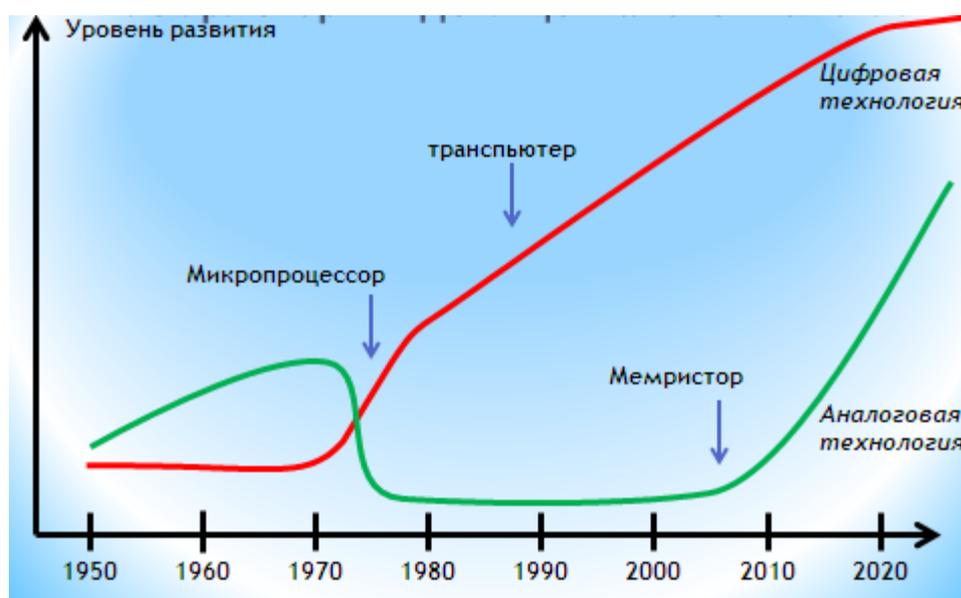


Рис.8. Развитие технологий нейрокомпьютеров

Необходимо отметить, что аналоговые ЭВМ в основном применялись для моделирования сложных динамических систем и управления ими, а также для решения систем уравнений в частных производных.

9. Особенности архитектуры современных и перспективных суперЭВМ

Основные архитектурные особенности современных и перспективных суперЭВМ проявились в процессе создания транспьютерных систем [19]. Такими свойствами были:

- масштабируемость;
- двухслойность архитектуры.

Масштабируемость определяла возможность практически линейного роста производительности при росте числа каскадируемых элементов (узлов) вычислительной системы (в некоторых определенных пределах).

Двухслойность как эффективное свойство архитектуры появилась в связи с появлением в конце 80-ых - начале 90-ых годов ограничений чисто транспьютерных (решетчатых) архитектур и ограниченной скоростью передачи данных между транспьютерами. Вычислительные системы строились тогда двухслойными, где первый, внутренний слой был чисто транспьютерным, а второй (внешний) слой обрабатывающей системы состоял из периферийных процессоров JMS A100 (специализированных на обработке сигналов), JMS A110 (специализированных на обработке изображений), JMS 860 (скалярных).

Эти два свойства эффективно проявились при создании в 2005-2007г.г. первых образцов суперЭВМ на базе графических процессоров фирмы nVidia. И эти два свойства должны в обязательном порядке сохраниться в образцах суперЭВМ, разрабатываемых в ближайшее время.

Однако, с нашей точки зрения, для перехода на экзафлопсный уровень вычислений необходима реализация по крайней мере еще одного шага – изменения логического базиса алгоритмов решения задач и соответственно элементной базы с переходом от носителя информации в виде уровней токов и напряжений в электрических схемах к носителю информации в виде частоты узких импульсов, подобно тому, как это имеет место в реальной нервной системе.

В одной из публикаций в прессе (Высокопроизводительные вычисления (мировой рынок) - TADVISER) от 24.11.2011г. отмечается следующее. «Белый дом заявил о нежелании втягиваться в гонку вооружений в вопросе создания все более быстрых компьютеров, и год назад в своем докладе предупредил, что «акцент на скорость может отвлечь ресурсы от фундаментальных исследований, направленных на развитие принципиально новых подходов, которые помогли бы нам, в конечном итоге, обогнать другие страны».

По давней традиции соредакторами ТОП500 суперЭВМ являются четыре человека – два из Америки и два из Европы. Один из соредакторов ТОП 500 с американской стороны Хорст Саймон (Horst Simon) дважды выступал на конференциях в мае 2013г. с программным докладом «Вычисления экза-масштаба и почему мы их не получим к 2020 году». На протяжении всей истории суперкомпьютеров рост производительности происходил с соблюдением (а иногда и некоторым опережением) известного закона Мура, согласно которому быстродействие вычислений повышается в 1000 раз примерно каждые 100-11 лет. Петафлопсный рубеж (10^{15} операций с плавающей запятой в секунду) был преодолен, как и планировалось, в 2008 г. Поэтому в суперкомпьютерном обществе фактически по умолчанию принято рассматривать 2018-2020 годы как выход на следующий, экзафлопсный, уровень вычислений (1 экзафлопс = 10^{18} операций в секунду). По мнению такого авторитетного эксперта как Хорст Саймон из всех существующих сейчас направлений развития суперЭВМ наиболее перспективным на ближайшие годы является направление суперЭВМ на базе графических процессоров. Но в этом направлении на пути к экзафлопсным вычислениям он видит ряд определенных трудностей, что позволяет ему заявить, что к 2020 году экзафлопсная суперЭВМ не будет создана.

В [35] представлены основные пути развития суперкомпьютеров в сторону экзафлопсного диапазона вычислений, где, на наш взгляд, мемристоры должны занять достойное место.

Основными предпосылками принципиального изменения архитектуры вычислительных систем при переходе к системам экзафлопной производительности являются:

- Необходимость резкого повышения надежности за счет отказа от фон-Неймановской архитектуры вычислительных систем из элементов булевой логики И, ИЛИ, НЕ, когда в системе из значительного числа этих элементов происходит катастрофический отказ при отказе любого элемента;

- Необходимость резкого снижения энергопотребления за счет отказа от существующего позиционного метода представления информации и перехода к другим, обеспечивающим снижение энергопотребления и как следствие, дополнительное повышение надежности;

- Необходимость резкого повышения однородности схемотехники элементов вычислительных систем, что должно привести к повышению эффективности при том резком увеличении интеграции элементов, которое будет иметь место при переходе к экзафлопной производительности.

10. Научно-технический задел по созданию отечественных супернейрокомпьютеров

Научно-техническим заделом, необходимым для разработки в России перспективных мемристорных высокопроизводительных вычислительных систем являются:

- Теория нейронных сетей, как методика синтеза структур из нейронов различного вида, алгоритмов адаптации весовых коэффициентов в этих структурах в процессе решения различных задач [3, 4, 5];

- Нейроматематика, как раздел вычислительной математики, связанный с решением в нейросетевом логическом базисе различных сложных формализуемых и неформализуемых задач [6];

- Нейроуправление, как раздел теории управления, связанный с применением нейрокомпьютеров в качестве систем идентификации сложных динамических систем и нейрокомпьютеров [7];

- Нейрокомпьютеры и нейрочипы [8].

Опыт российских ученых в этих областях науки обобщен в монографии [28].

Уровень развития микроэлектроники в 60ых и 70ых годах прошлого столетия позволял возникнуть идее построения адаптивных элементов с аналоговой памятью. Именно поэтому необходим детальный анализ (конечно с

позиций развития современного уровня технологии микроэлектроники – нанотехнологии) разработок адаптивных элементов с аналоговой памятью прежних лет, даже 60ых и 70ых годов прошлого столетия [36, 37, 38].

Требуют детального изучения многочисленные разработки аналоговых и аналого-цифровых нейрочипов последних десятилетий с целью использования в современных разработках мемристорных систем, в том числе:

- аналоговые нейрочипы;
- аналого-цифровые нейрочипы;
- клеточные нейрочипы;
- нейрочипы с частотно-импульсным представлением сигналов;
- оптические и оптоэлектронные нейрочипы;
- молекулярные нейрочипы;
- специализированные аналоговые и аналого-цифровые нейрочипы;
 - АЦП;
 - СМАС;
 - обработки изображений;
 - нейроруправления;
 - ассоциативной памяти;
 - обработки речевой информации;
 - другие.

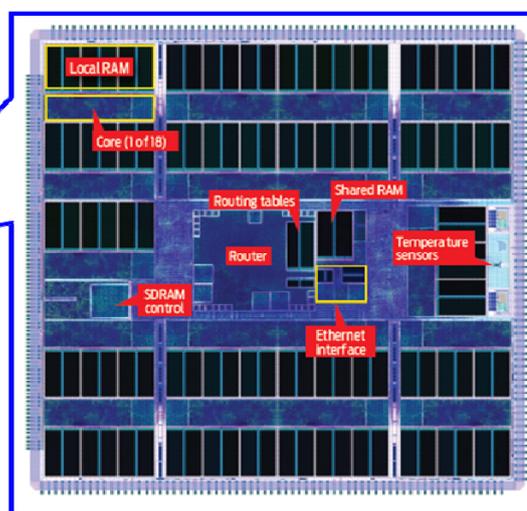
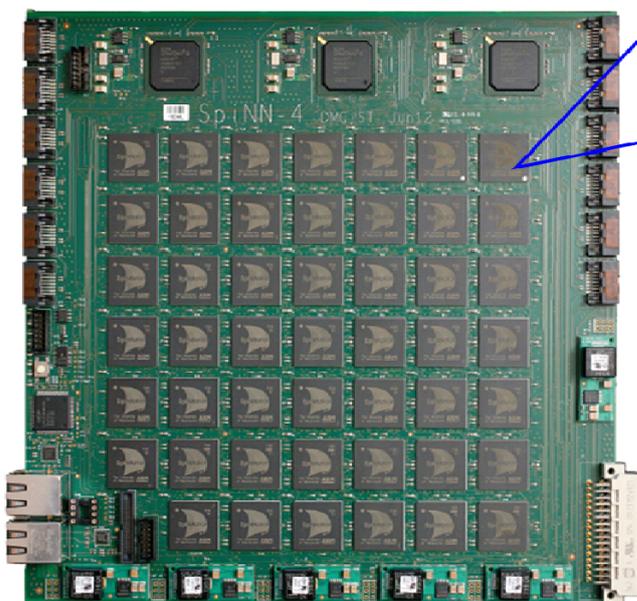
Следует отметить необходимость анализа двух новых программ: Brain Initiative (США) и Human Brain (ЕИ), объединяющих усилия многих организаций и университетов США и Европы, а также некоторых текущих и прошедших работ в области моделей мозга, нейрокомпьютеров и соответствующих прикладных задач:

- Проект Irvine Sensors;
- Проект Ливерморской лаборатории;
- Проект Института продвинутых архитектур (Сандиа, Оак Ридж);
- Проект Интел, центр суперкомпьютерных вычислений Сан-Диего, DARPA;
- Проект eLiza самоуправляемой автономной компьютерной системы (IBM);
- Проект SyNAPSE (Systems Neuromorphic Adaptive Plastic Scalable Electronics);
- Проект Neurogrid (Стэнфордский университет);

- Проект SpiNNaker (Spiking Neural Network Architecture) – группа английских университетов;
- BICA (Biologically Inspired Computer Architecture);
- Проект BlueBrain;
- MONETA (Modular Neural Exploring Traveling Agent);
- Cog-Ex-Machina;
- Marble;
- МНАНН;
- Neu-Weu;
- Проект нейроморфного компьютера с использованием мемристоров (Интел совместно с МТИ);
- проект SPAUN.

На рис.9 представлена типичная передовая разработка супернейрокомпьютера (базовая плата и специализированный микропроцессор) SPiNNaker для решения сложных вычислительных задач.

**Базовая плата
перспективного
супернейрокомпьютера
для расчетных задач**



**Структура базовой
СБИС перспективного
супернейрокомпьютера
для расчетных задач**

Рис.9 Разработка супернейрокомпьютер SPiNNaker для решения сложных вычислительных задач

11. Мемристоры – детище нанотехнологий

На рис. 10 условно представлено развитие технологий нейрокомпьютеров на фоне развития вычислительной техники. Отмечена динамика развития цифровой и аналоговой вычислительной техники.

В части **цифровой вычислительной техники** отмечены:

- постепенный рост производительности в период 50ых-60ых годов прошлого столетия;
- интенсификация развития в начале 70ых годов прошлого столетия при появлении микропроцессора;
- революционное развитие с начала 80ых годов прошлого столетия вследствие появления транспьютера;
- затухание развития в период 2010-2020 г.г. вследствие технологических ограничений.

В части **аналоговой вычислительной техники** отмечены:

- рост мощности аналоговых моделирующих комплексов в период 50ых-60ых годов прошлого столетия;
- резкий спад развития вследствие появления микропроцессора;
- отсутствие активного развития при появлении транспьютера;
- резкий рост развития вследствие развития нанотехнологий и появления мемристора и мемристорных систем.



Рис. 10 Развитие технологий нейрокомпьютеров на фоне развития вычислительной техники

Мемристоры – новое кардинальное направление развития элементной базы перспективных высокопроизводительных вычислительных систем.

На рис.11 представлено место мемристоров в общем комплексе элементов, составляющих схемы электронных устройств.

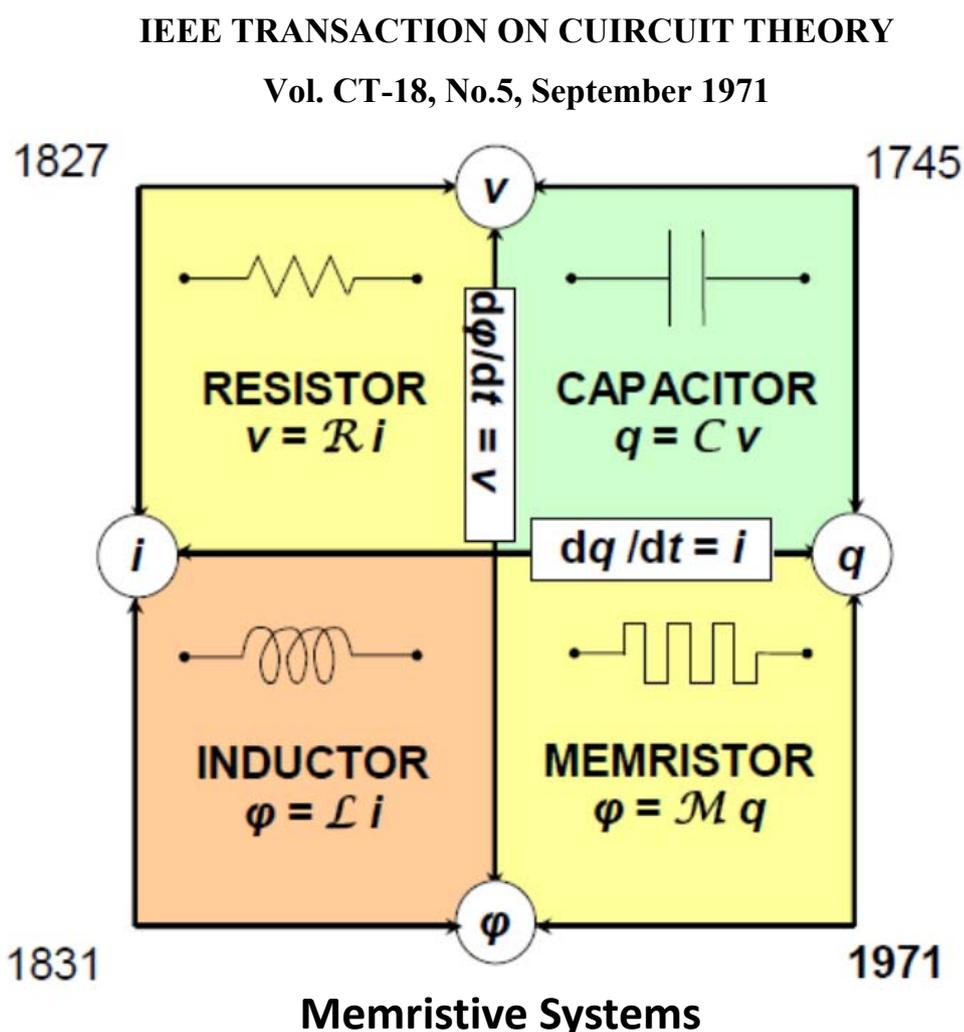


Рис. 11. Место мемристоров в общем комплексе элементов, составляющих схемы электронных устройств

На рис.12 представлен вариант построения электрической схемы фрагмента нейронной сети с комплексированием мемристоров и КМОП-схем.

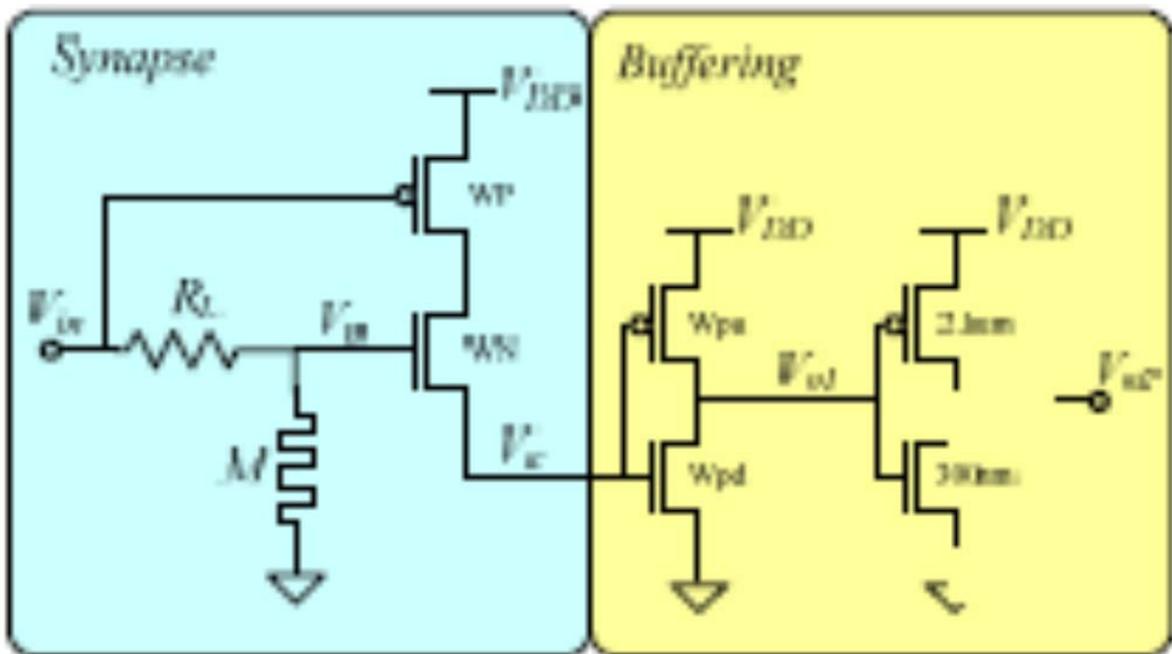


Рис. 12. Вариант построения электрической схемы фрагмента нейронной сети с комплексированием мемристоров и КМОП-схем

Другим принципиальным свойством мемристорных систем является переход к представлению информации в системе в виде частоты узких импульсов, что позволяет резко снизить потребление энергии и, как следствие, увеличить надежность вычислительной системы.

Разработка мемристоров требует создания их моделей для решения задач их проектирования и использования в дальнейшем при создании систем на мемристорах. Модели могут быть разработаны в следующих вариантах:

- аналитические модели,
- модели на GPU,
- модели на базе Matlab,
- модели на базе Spice,
- модели на FPGA,
- модели в Cadence,
- и др.

необходимо отметить, что модели мемристорных систем на GPU и FPGA могут использоваться как самостоятельные изделия современной высокопроизводительной вычислительной техники.

12. Первоочередные работы в области мемристорных систем

Учет российского опыта в области разработки нейрокомпьютеров позволяет сформулировать перечень первоочередных работ в части исследования систем на мемристорах:

1. Разработка схем и принципов стыковки мемристорных сетей с КМОП схемами.

2. Разработка инструментальной системы для исследования свойств мемристоров с выходом на ПЭВМ, в том числе контроллера на ПЛИС для стыковки мемристорных систем с персональным ЭВМ, с драйвером.

3. Экспериментальная и теоретическая оценка понижения энергопотребления в мемристорных системах при переходе к представлению сигналов в виде частотно-модулированной последовательности узких импульсов.

В число первоочередных работ в области мемристорных систем должны входить разработка архитектуры и технических предложений по реализации базовых нейросетевых систем на мемристорах:

- бинарные ОНС;
- каскадируемый фрагмент слоя многослойной нейронной сети;
- каскадируемый фрагмент СМАС-структуры;
- каскадируемый фрагмент клеточной нейронной сети;
- фрагмент хаотической нейронной сети;
- каскадируемый фрагмент нейронной сети с обратными связями;
- ассоциативная память;
- 3D-мемристорные системы.

13. Анализ технологий изготовления мемристоров и мемристорных систем

Необходимо отметить некоторое множество технологий изготовления мемристоров, например:

- металл-оксид;
- оксид титана;
- оксид тантала;
- спинтроника;
- полимеры;

- ферроэлектрики;
- углеродные нанотрубки;
- кремний;
- некристаллический кремний;
- аморфный кремний;
- поликремний;
- электрохимия;
- барьер Шоттки.

При этом задачей разработчика вычислительной системы является анализ и оценка выбираемой технологии по следующим двум признакам:

1. Признаки ограничений технологии на топологию и характеристики реализуемого фрагмента нейронной сети для будущего использования этих ограничений при разработке алгоритмов настройки нейронных сетей и алгоритмов решения задач.

2. Признаки ограничений технологии на показатели производительности будущего нейрокомпьютера:

- число эмулируемых нейронов;
- число эмулируемых настраиваемых связей;
- число переключений связей в единицу времени;
- энергопотребление.

Расчет по второму признаку необходимо проводить на уровне:

- элементарных операций;
- базовых нейросетевых систем на мемристорах;
- СБИС;
- плат;
- блоков;
- супернейрокомпьютера;

с учетом архитектуры и схемотехники коммутационной среды каждого элемента.

Отдельно необходимы анализ и сравнение реализации нейросетевых систем на GPU и мемристорах.

14. Замечание к оценке производительности

Подобно косвенным показателям технической производительности вычислительных систем с фон-неймановской архитектурой (MFlops, TFlops) для нейροкомпьютеров существуют подобные косвенные показатели производительности, определяемые следующим образом:

- число эмулируемых нейронов;
- число эмулируемых настраиваемых связей;
- число переключений связей в единицу времени.

Простая техническая оценка времени выполнения мемристорами и мемристорными системами простейших операций является достаточно эффективной. Однако нужно всегда помнить историю развития оптических и оптоэлектронных компьютеров, в которых первоначальной идеей было резкое ускорение некоторых матричных операций при реализации их на оптических элементах. При этом реализация других операций (кроме матричных) в алгоритмах решения задач потребовало обычных цифровых схем текущего уровня развития технологии микроэлектроники. В совокупности подобная оптоэлектронная вычислительная система оказалась малоэффективной в ориентации на широкий круг прикладных задач. Причина этого заключается как раз в том, что матричные операции, эффективно выполняемые на оптических элементах, являются только частью общего алгоритма решения задачи.

В свое время была предпринята попытка реализации нейροкомпьютеров на оптических элементах [9], вследствие большой важности матричных операций в реализации нейросетевых алгоритмов. Эта попытка также не увенчалась успехом, т.к. в то время трудно было реализовать в оптическом диапазоне массовые нелинейные операции (реализация функций активации нейронов), а также алгоритмы настройки весовых коэффициентов. Попытки реализации этих нелинейных блоков использованием оптических резонаторов также не увенчались успехом.

Подобная история повторилась с транспьютерными сигнальными процессорами JMS A100, реализующими каскадируемый Z-фильтр 32 порядка с настраиваемыми коэффициентами с переменной, управляемой пользователем, разрядностью. 3 наносекунды на операцию умножения-сложения не сделала данные СБИС основой универсальных ЭВМ, т.к. в виде Z-фильтра представлялась

лишь часть, иногда весьма незначительная, общего алгоритма решения задачи. Именно поэтому СБИС JMS A100 были эффективными только для задач обработки сигналов, а СБИС JMS A110 для задач обработки изображений [19].

Развитие нейроматематики за последние 20-25 лет [6] привело к тому, что алгоритмы решения большинства, если не всех сложных формализуемых и неформализуемых задач могут быть представлены практически полностью в виде нейросетевых структур различного вида. Это делает мемристорные системы потенциально эффективными для использования в разработке как специализированных, так и универсальных нейрокомпьютеров.

Необходимо отметить, что как и для обычных вычислительных машин, для нейрокомпьютеров существует более объективная, чем косвенная, оценка производительности – время решения задачи при заданной точности и реализуемом энергопотреблении.

15. Сфера применения будущих мемристорных систем

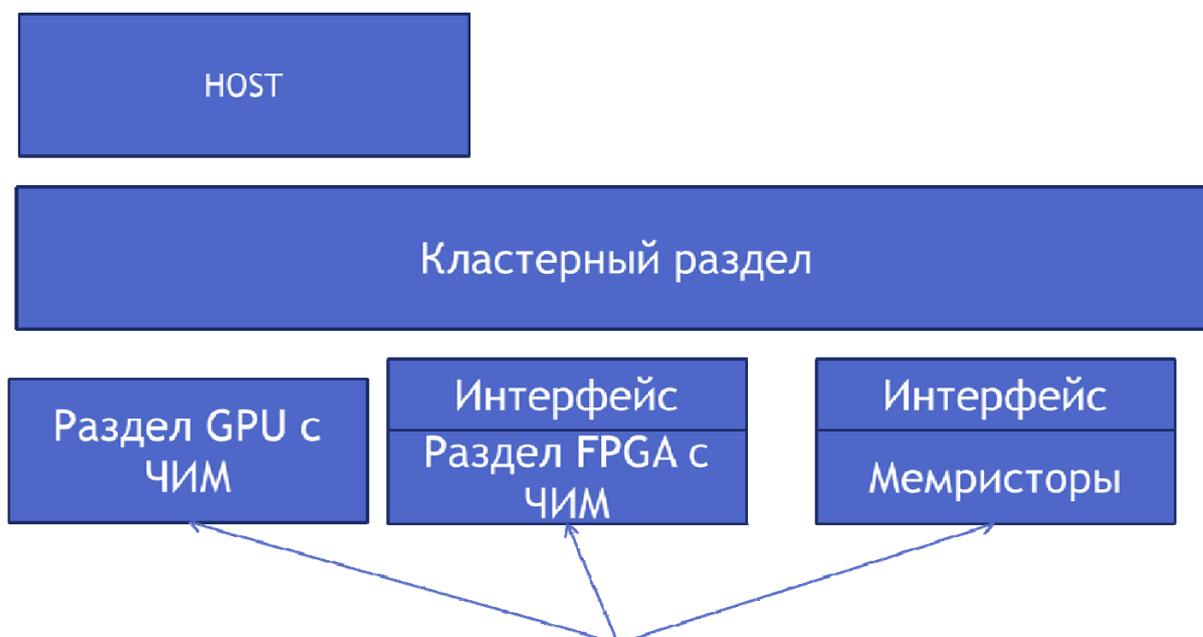
Наверняка мемристоры и мемристорные системы найдут самое широкое применение. Сейчас, на начальном этапе их разработок, применения сосредотачиваются около следующих направлений:

- коммутаторы;
- память;
- RFID-системы;
- распределенные сенсорные системы;
- нейропротезы;
- видеоаналитика;
- робототехника;
- обработка сигналов;
- системы защиты информации.

Необходимо отметить также одно из важнейших перспективных применений мемристоров – реализация супернейрокомпьютеров [29, 33].

16. Архитектура супернейрокомпьютера с использованием мемристоров

Архитектура супернейрокомпьютера с использованием мемристоров представлена на рис.13.



Оптимальное распределение фрагментов решаемой задачи

Рис. 13 Архитектура супернейрокомпьютера с использованием мемристоров

Ядром данной вычислительной системы является классическая кластерная архитектура, использующая современные микропроцессоры. Второй слой обрабатывающих элементов представлен тремя вариантами:

- GPU;
- FPGA;
- Мемристоры.

В каждом из этих блоков эмулируется или реализуется частотно-модулируемая последовательность импульсов, как основной источник информации в реализуемых нейронных сетях. Прикладной программист оптимально распределяет фрагменты решаемой задачи между тремя типами сопроцессоров в соответствии с их свойствами.

17. Базовые технические средства

В состав работ первой очереди по созданию базовых технических средств супернейрокомпьютеров входят следующие:

- Разработка архитектуры и экспериментальных образцов каскадируемых СБИС на базе мемристорных матриц для использования в перспективных супернейрокомпьютерах.

- Разработка архитектуры и экспериментальных образцов базовых каскадируемых плат в конструктивах PCI, microPCI, VME, ... (типа Spinnaker) с использованием СБИС на базе мемристоров для использования в перспективных супернейрокомпьютерах.

- Разработка архитектуры и экспериментальных образцов базовых каскадируемых блоков с использованием плат в конструктивах PCI, microPCI, VME, ... для использования в перспективных супернейрокомпьютерах.

- Разработка ассоциативной памяти.

- Оформление патентов на архитектуры СБИС, плат и блоков на базе мемристорных матриц.

18. Системное программное обеспечение

В состав системного программного обеспечения супернейрокомпьютера должны входить, по крайней мере, следующие блоки:

- общее системное программное обеспечение (ОС, компиляторы и пр.);

- алгоритмы и программы распараллеливания нейронных сетей различной структуры;

- вариант MPI;

- вариант CUDA;

- алгоритмы и программы настройки нейронных сетей с учетом ограничений мемристоров.

последний раздел должен быть достаточно полным развитием методов настройки многослойных нейронных сетей, изложенных в работах [4, 5] с учетом реальных динамических характеристик мемристоров и мемристорных систем при реализации фрагментов алгоритмов решаемых задач на аналоговой части супернейрокомпьютера. Это нужно сделать подобно тому, как в [4, 5] разработанные алгоритмы настройки нейронных сетей были модифицированы для ограничений на весовые коэффициенты типа равенств или неравенств, присутствующих в аналоговых реализациях нейрокомпьютеров 60ых и 70ых годов прошлого столетия [8]. Для оценки качества разрабатываемых алгоритмов настройки нейронных сетей должно быть создано тестовое программное обеспечение, по крайней мере для режимов обучения и самообучения, с тестами в виде генераторов случайных сигналов с многомодальным распределением и

различными модальностью, размерностью пространства признаков и степенью пересеканости мод функции распределения.

19. Оценка качества вычислительных систем с использованием мемристоров

Данный раздел работ по созданию супернейрокомпьютеров с использованием мемристоров должен содержать следующее:

1. Разработку и апробацию методики количественной оценки производительности с учетом п.13;
2. Разработку методики обеспечения и оценки необходимой точности вычислений при переходе в супернейрокомпьютере к аналоговой обработке;
3. Оценку понижения энергопотребления в супернейрокомпьютере при переходе к представлению сигналов в виде частотно-модулированной последовательности узких импульсов.

20. Состав работ ОКР по созданию образцов супернейрокомпьютера на базе мемристоров (2016-2018 гг.)

В состав данных работ входят следующие:

1. Разработка технических средств супернейрокомпьютера на базе мемристоров.
 - 1.1. Разработка и изготовление опытной партии СБИС нейрочипов на базе мемристоров.
 - 1.2. Разработка и изготовление опытной партии нейроплат с использованием нейрочипов на базе мемристоров.
 - 1.2.1. Разработка архитектуры платы.
 - 1.3. Разработка и изготовление опытной партии блоков для размещения нейроплат с использованием нейрочипов на базе мемристоров.
2. Разработка системного программного обеспечения супернейрокомпьютера на базе мемристоров.
3. Комплексная отладка супернейрокомпьютера на базе мемристоров.
4. Проверка работоспособности и эффективности супернейрокомпьютера на прикладных задачах.

21. Разработка прикладного программного обеспечения супернейрокомпьютера с использованием мемристоров

В качестве базовых прикладных задач для реализации на супернейрокомпьютере с использованием мемристоров выбраны либо важные задачи государственного уровня, либо задачи, по которым в течение последних 20-25 лет сформирован большой научно-технический задел в нейроматематике.

Первой такой прикладной задачей является задача создания пакетов программ Neuron-GPU, Neuron-FPGA, Neuron-M решения, с использованием каждого из приведенных типов сопроцессоров, базовых аналитических задач в нейросетевом логическом базисе:

- экстраполяция функций;
- распознавание образов;
- кластеризация;
- возможно других.

При этом при выборе прикладных задач необходимо учитывать, что основными направлениями применения различных высокопроизводительных вычислительных систем являются следующие:

- биоинформатика,
- криптография,
- ядерные исследования,
- аэро-, гидродинамика,
- моделирование мозга.

В части биоинформатики, а именно проблем обработки геномной и протеомной информации, необходимо отметить, что, начиная с 2006г. значительное число научных работ, в основном по этой теме, в журнале IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) посвящено применению нейросетевых технологий.

В последние годы намечается тесный контакт НБИКС технологий и геномно- протеомных исследований. В явном виде эта связь обозначена в Европейском проекте Human Brain, где задачи исследования мозга:

- когнитивные проблемы;
- функционирование интегрального мозга;
- изучение связей между разделами мозга;
- изучение разделов мозга;

- клеточные образования;
- клетки;
- синапсы;

тесно связываются с проблемами исследования генома, транскриптома, протеома, метаболома, с надеждой на получение более точно функционального описания в решении задач исследований мозга. Целью подобных исследований может быть описание синапсов, клеток, и более сложных образований геномными, протеомными, полно транскриптомными, метаболомными и секретомными картограммами и профилями. Более подробно эти идеи изложены в проекте российской программы, разработанной под руководством проф., д.м.н. А.С. Брюховецкого, «Проект «Мозговой шторм»» (Brainstorming project) [32].

Активное развитие в последние годы применения нейросетевых технологий в решении различных задач криптографии отмечено в монографии [30].

Нейросетевые технологии активно применяются во всем мире в ядерных исследованиях. В первую очередь это относится к решению следующих задач:

- решение различных задач газодинамики;
- нейроуправление плазмой;
- различные задачи контроля и управления в АЭС;
- нейроуправление центрифугами;
- обработка инструментальной информации в исследовательских центрах.

Задачи аэро-, гидродинамики связаны с решением сложных 2D и 3D систем специальных дифференциальных уравнений в частных производных. Опыт российской научной школы в области применения нейросетевых технологий в решении этих задач отражен во многих научных работах за последние 10-15 лет, в том числе в нескольких монографиях и диссертациях (докторских и кандидатских).

Исследования в области моделирования мозга и его разделов в настоящее время проводятся во всем мире. Их результаты показывают реальные ограничения сверхвысокопроизводительной вычислительной техники на текущем уровне развития.

Именно поэтому в реализации прикладных задач на супернейрокомпьютере с использованием мемристоров выбраны следующие:

- решение задач нейрокриптографии;

- реализация нейросетевых алгоритмов обнаружения и отражения атак на информационные системы;
- решение задач биоинформатики;
- решение уравнений математической физики;
- решение задач управления плазмой;
- обработка больших данных (bigdata);
- создание 3D моделей мозга.

22. Нейропротезы – эффективное направление применения НБИКС технологий

В последние годы искусственные протезы фрагментов нервной системы находят все более и более широкое применение. Разработки подобных протезов имеют как бы двойное применение. С одной стороны эти разработки имеют чисто медицинское направление, ориентированное на помощь инвалидам, потерявшим зрение, слух, конечности, возможность двигаться и т.д. С другой стороны эти разработки необходимы для реализации различных робототехнических систем [27].

В настоящее время нейропротезы реализуются для помощи инвалидам в следующих направлениях:

- зрительный имплантат для помощи незрячим больным;
- слуховой имплантат для помощи слабослышащим;
- нейроимплантат управления протезами кистей и рук;
- нейроимплантат некоторых разделов мозга.

Здесь при создании нейропротезов значительную роль играют нейросетевые технологии.

Необходимо отметить, что в настоящее время человечество в этой области находится на самой начальной стадии развития этой проблемы, затронув только внешние, периферийные системы и находится далеко от настоящего понимания принципов функционирования мозга, несмотря на реальные попытки воспроизведения некоторых высших функций нервной системы (распознавание, прогнозирование, управление движением и т.п.).

23. О роли нейрофизиологических исследований

Нейрофизиологические исследования с позиций инженера можно разделить на две части:

- направленные на решение медицинских задач;
- направленные на понимание структуры и принципов работы нервной системы и ее разделов с целью возможного воспроизведения.

В проблеме НБИКС нейрофизиологические исследования тесно связаны с разработкой перспективных сверхвысокопроизводительных вычислительных систем, в том числе сверхвысокопроизводительных систем с ограниченными габаритами и энергопотреблением.

Зачастую сверхвысокопроизводительные вычислительные системы с ограничениями по весам, габаритам и энергопотреблению не могут в принципе быть реализованы на элементной базе широкого применения. Это касается таких применений как малогабаритные роботы, медицинские роботы, нейропротезы и т.п.

Нейрофизиологические исследования нервной системы (в том числе мозга) человека и животных развиваются во времени параллельно развитию технологии микроэлектроники. Эти два направления взаимно обогащают и питают друг друга (рис. 14).

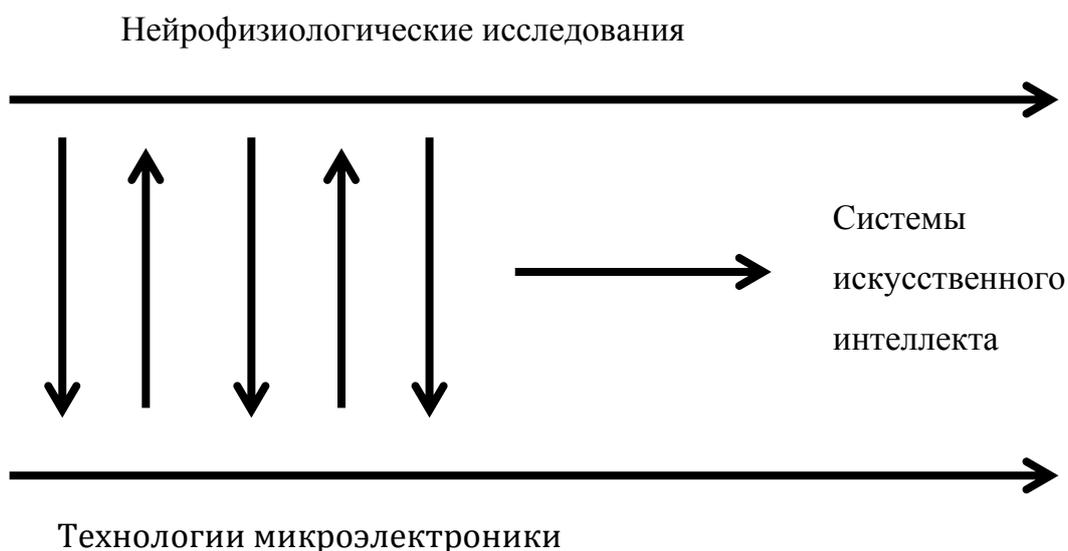


Рис. 14. Условная схема взаимодействия направлений нейрофизиологических исследований и технологий микроэлектроники.

Со временем с развитием технологий микроэлектроники, в частности с развитием нанотехнологий, связь этих двух направлений развития науки все более и более углубляется. Конечной целью этих работ в будущем будут эффективные системы искусственного интеллекта.

В заключение данного раздела отметим основные направления использования нейрофизиологических исследований при создании высокопроизводительной вычислительной техники:

- нейроимплантанты (глаз, ухо, ...);
- вычислительные системы управления функциями гуманоидных роботов;
- интерфейс «мозг-компьютер»;
- новые гипотезы построения нейронных сетей для решения задач.

24. НБИКС технологии и проблема искусственного интеллекта

Основная первопричина появления и расширения задач искусственного интеллекта (ИИ) это развитие самого общества, связанные с ним развитие технологий и появление новых трудных задач для решения.

Понятие «искусственный интеллект» (artificial intelligence) в последние годы заменено на понятие «интеллектуальные вычисления» (intelligence computational). Всеми задачами искусственного интеллекта занимается международное общество IEEE Intelligence computational society.

В принципе можно считать, что проблема искусственного интеллекта является искусственно надуманной. По сути дела, есть:

- сложные неформализуемые и слабоформализуемые задачи;
- различные технологии для их решения;
- различные методы реализации;
- нейрофизиологические исследования;

На рис.15 условно представлена динамика развития во времени сложности решаемых с помощью вычислительных систем задач:

- для классических методов вычислительной математики;
- для нейросетевых методов решения задач.

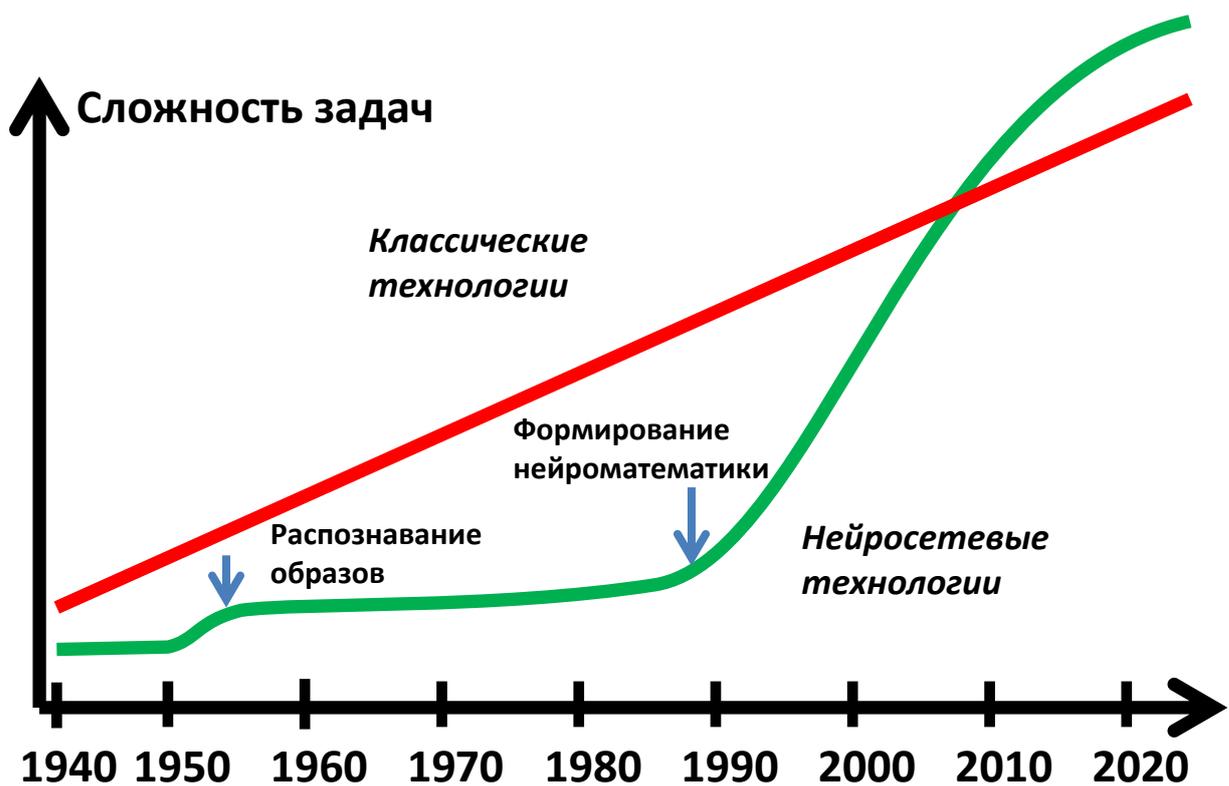


Рис.15 Развитие технологий решения задач

На рис. 15 отмечены три качественных момента развития нейросетевых технологий:

- в начале 50ых годов прошлого столетия с появлением персептронов;
- в начале 90ых годов прошлого столетия с началом активного формирования и развития нейроматематики;
- развитие нейроматематики после 2000 года показало, что нейросетевые алгоритмы более пригодны для решения сложных задач, чем классические.

Особенностью решения слабоформализуемых и неформализуемых задач, преимущественно использующих для своего решения нейросетевые технологии, является то, что заранее нельзя предсказать или задать качество решения задачи. Повышение требований к качеству решения неформализуемой задачи может потребовать столь значительного увеличения мощности алгоритма и, как следствие, объема вычислительной системы при ограничениях на время решения задачи. Может случиться так, что эти требования приведут к физической нереализуемости вычислительной системы при заданных ограничениях на объемовесовые характеристики вычислительной системы.

Технологии решения задач можно подразделить следующим образом:

- классические математические;
- нейросетевые технологии;
- технологии нечеткой логики;
- эволюционное программирование.

С этой точки зрения в состав задач искусственного интеллекта можно включить:

- только неформализуемые задачи;
- все вышеперечисленные технологии кроме классических математических.

Методы реализации задач ИИ имеют весьма широкий спектр:

- от программной реализации на персональных ЭВМ;
- до разработки специальных нейрочипов.

До появления масштабируемой вычислительной техники всякие разработки в области ИИ практически до 90ых годов прошлого столетия вызвали раздражение по той причине, что разработчики систем ИИ для реализации своих идей брали те вычислительные системы, которые были у них под рукой (ЕС ЭВМ, СМ ЭВМ, «Электроника-60» и др.), не имея возможности увеличить вычислительную мощность при необходимости сокращения времени решения задачи или увеличения точности решения.

Основная проблема в том, что, как правило, для решения задач ИИ используют вычислительные системы уровня текущего дня. При этом оказывается, что использование вычислительной машины заданной мощности (производительности) не может в принципе обеспечить заданное или требуемое качество решения задачи или решить задачу ИИ в требуемое время.

В этом смысле, с одной стороны, развитие проблемы ИИ в значительной степени есть развитие вычислительной техники.

С другой стороны, развитие вычислительной техники способствует расширению класса задач ИИ, решаемых на практике.

С точки зрения вышеизложенного первоочередными задачами ИИ на настоящий момент можно считать следующие:

- разработки супернейрокомпьютеров, в том числе с использованием отечественной элементной базы,

- информационная безопасность, включая нейросетевые системы обнаружения и отражения атак в информационных системах,
- нейрокриптография,
- нейросетевые технологии в геномных и протеомных исследованиях,
- моделирование разделов мозга,
- нейропротезы.

Вышеизложенное позволяет определить понятие «искусственный интеллект» (по крайней мере, на настоящее время) так, как это представлено на рис. 16.

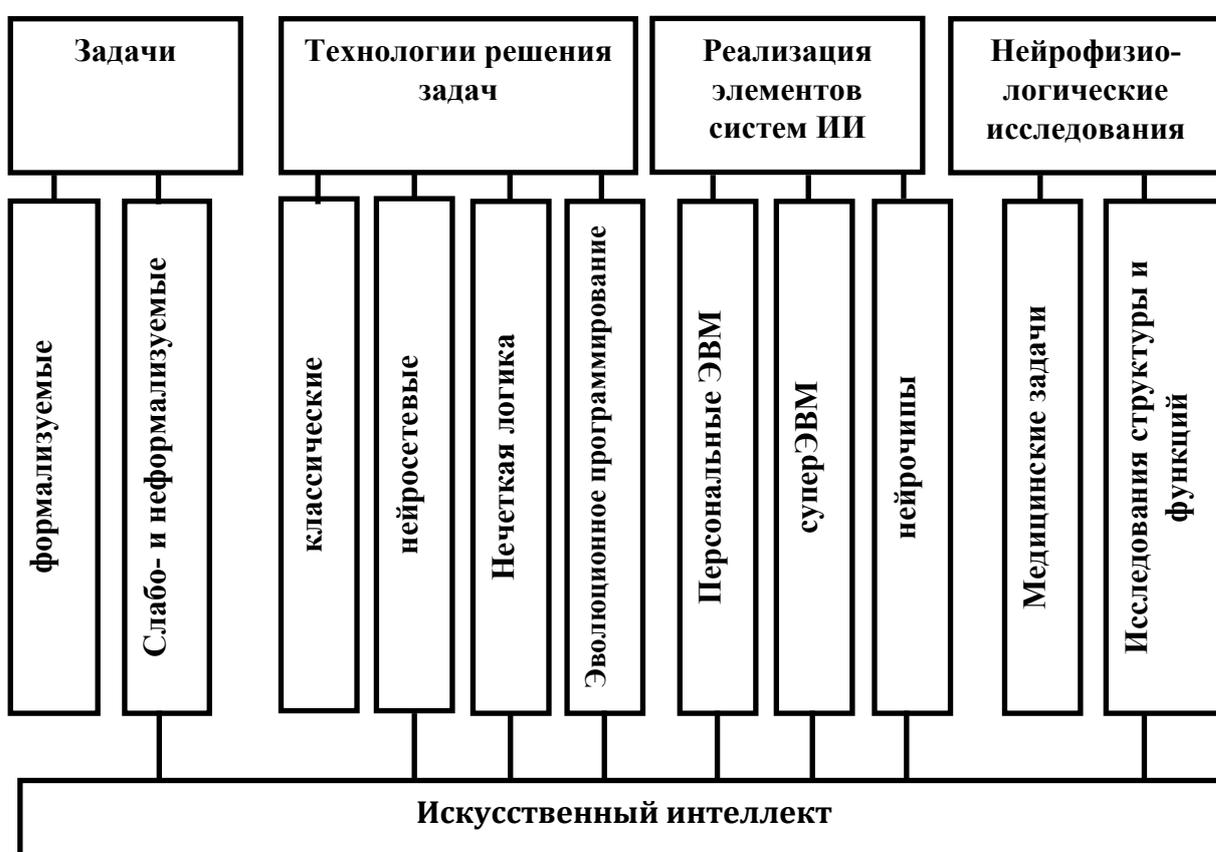


Рис. 16. К определению понятия «искусственный интеллект» (ИИ) (по крайней мере, на настоящее время)

Заключение

1. Генеральной линией развития НБИКС технологий является максимально возможное в обозримый период времени изучение принципов функционирования нервной системы, включая мозг, человека и животных, попытки эффективной реализации и применения нейросетевых технологий.

2. Нанотехнологии привели к возможности разработки микросхем очень большой интеграции и, как следствие, к развитию нейросетевых вычислительных систем, являющихся естественным продолжением развития сверхвысокопроизводительной вычислительной техники.

3. Биологические исследования активно используют и будут использовать нейросетевые технологии, по крайней мере в двух направлениях:

- в части нейрофизиологических исследований мозга и нервной системы человека и животных, в частности в разработке различных нейропротезов;
- в части обработки геномной и протеомной информации.

4. Информационные технологии выходят на уровень активного использования нейросетевых технологий в части развития:

- теории нейронных сетей;
- нейроматематики;
- нейроуправления;
- нейрокомпьютеров и нейрочипов.

5. Когнитивные исследования в основном связаны с решением сложных неформализуемых задач, для которых в свою очередь основными для решения являются нейросетевые технологии.

6. Социальные исследования являются зачастую сложными неформализуемыми и слабоформализуемыми задачами:

- аналитические задачи в системах государственного управления, включая системы принятия решений и экспертные системы;
- модели социальных систем;
- задачи прогнозирования социальных явлений.

В решении указанных задач значительное место занимают нейросетевые технологии.

7. Дальней целью представленных исследований является создание искусственного мозга. Необходимо отметить, что НБИКС технологии являются практически единственным, основным направлением на пути решения этой проблемы. Мемристорные системы – очередной шаг на пути развития вычислительной техники в реализации этой проблемы. За ним должны последовать следующие шаги, определяемые развитием нанотехнологий.

Литература

1. Маккалок У.С., Питтс У.В. Логическое исчисление идей, относящихся к нервной активности. «Нейронные сети: история развития теории», // книга 5 серии «Нейрокомпьютеры и их применение», Издательство Радиотехника, М., 2001г.
2. Kim K. H, Gaba S., Wheeler D., Cruz-Albrecht J. M., Hussain T., Srinivasa N., Lu W. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications // Nano Letters, 12, 2012.
3. Розенблатт Ф. Принципы нейродинамики. // Издательство Мир, 1964г.
4. Galushkin A.I. Neural Network Theory // Springer, 2007
5. Галушкин А.И. Нейронные сети: основы теории // Издательство «Горячая линия – Телеком», М., 2010г
6. Нейроматематика, серия «Нейрокомпьютеры и их применение». М., 2002. Т.6;
7. Галушкин А.И. Основы нейроуправления. // Информационные технологии. Приложение к журналу. 2002. №10.
8. Галушкин А.И. Нейрокомпьютеры. М., 2000;
9. Optical neural networks. Wikipedia;
10. Квантовые нейронные сети. МИФИ, М., 2001;
11. Рамбиди Н.Г. Биомолекулярные нейросетевые устройства, серия «Нейрокомпьютеры и их применение», кн. 33. М., Изд-во ИПРЖР, 2002.
12. I.G. Delgado-Frias, W.R. Moore. VLSI for Artificial Intelligence, Kluwer Academic Publ., 1989;
13. Elmasry M.I. (ed) VLSI Artificial neural networks engineering, Kluwer Academic Publ., 1994;
14. A.I. Annema. Feed-Forward neural networks. Vector decomposition, analysis, modeling a analog implementation. Kluwer Academic Publ., 1995;
15. Learning on silicon Adaptive VLSI Neural Systems. Kluwer Academic Publ., 1999.
16. Tsutomu Miki (ed) Brainware: Bio-Inspired Architecture and its Hardware Implementation World Scientific. 2001;
17. D. Zhang, S.K. Pal (ed) Neural Networks and Systolic Array design. World Scientific. 2002;

18. Omondi A.R., Rajapakse I.C. (ed) FPGA Implementations of Neural Networks. Springer, 2006.
19. Галушкин А.И., Точенов В.А. Транспьютерные системы – начало становления в России ЭВМ с массовым параллелизмом // Нейрокомпьютер, №3, 2005г.
20. Б.Б. Смоллов, Е.П. Угрюмов Время-импульсные вычислительные устройства // Издательство «Радио и связь», М., 1983
21. Strukov D. B., Snider G. S., Steward D. R., Williams R. S. The missing memristor found // Nature, 453, 2008.
22. Jo S. H., Chang T., Ebong I., Bhadviya B. B., Mazumder P., Lu W. Nanoscale memristor device as synapse in neuromorphic systems // Nano Letters, 10, 2010.
23. Kim H., Sah M. P., Yang C., Roska T., Chua L. O. Memristor bridge synapses // Proceedings of the IEEE, 100, 2012.
24. Yang J. J., Strukov D. B., Stewart D. R. Memristive devices for computing // Nature Nanotechnology, 8, 2013.
25. Strukov D.B., Stewart D.R., Borghetti J., Li X. Hybrid CMOS/memristor circuits // Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium.
26. Sharad M., Augustine C., Panagopoulos G., Roy K. Proposal for neuromorphic hardware using spin devices // Disordered Systems and Neural Networks, 2012.
27. Галушкин А.И., Занавескин М.Л. Нейрокомпьютер – вызов НБИКС-технологии в XXI веке // Природа, (в печати)
28. Галушкин А.И., Симоров С.Н. Нейросетевые технологии в России (1982-2010 гг.) // Издательство «Горячая линия – Телеком», М., 2012г.
29. Галушкин А.И. Стратегия развития современных супернейрокомпьютеров на пути экзафлопных вычислений // Приложение к журналу «Информационные технологии», №3, 2012г.
30. Червяков Н.И., Евдокимов А.А., Галушкин А.И., Лавриенко И.Н., Лавриенко А.В. Применение искусственных нейронных сетей и систем остаточных классов в криптографии // М., Физматлит, 2012г.
31. Галушкин А.И. Нейросетевые технологии в реставрологии // Информационные технологии, №9, 2011г.

32. Брюховецкий А.С. Проект «Мозговой штурм» (Brainstorming) (рабочий вариант) // ФГБУ Научно-клинический центр специализированных видов медицинской помощи и медицинских технологий ФМБА России. Клиника восстановительной интервенционной неврологии и терапии («Нейровита»).

33. Экзафлопные нейросетевые технологии. Концепция развития технологий высокопроизводительных вычислений на базе супернейрокомпьютеров (2012-2020гг.), 2012г.

34. Разработка прикладных технологий для инженерных задач с плотными системами уравнений и сверхбольшим числом неизвестных. // Госконтракт ФАНИ. Шифр ИТ-13а3/001, 2006г.

35. Берд Киви. К точке критического перехода // 3D-News (Daily Digital Digest), 29.05.2013

36. Аналоговые запоминающие и адаптивные элементы (под ред. Б.С.Состкова), Из-во «Энергия», М. 1973

37. В.В.Трейер, А.Б.Елизаров. Электрохимические интегрирующие и аналоговые запоминающие элементы. Из-во «Энергия», М. 1973

38. Галушкин А.И., Юмашев С.Г. Адаптивные элементы с аналоговой памятью. (рукопись неопубликованной монографии, представленной в Из-во «Энергия» в 1969г.)