



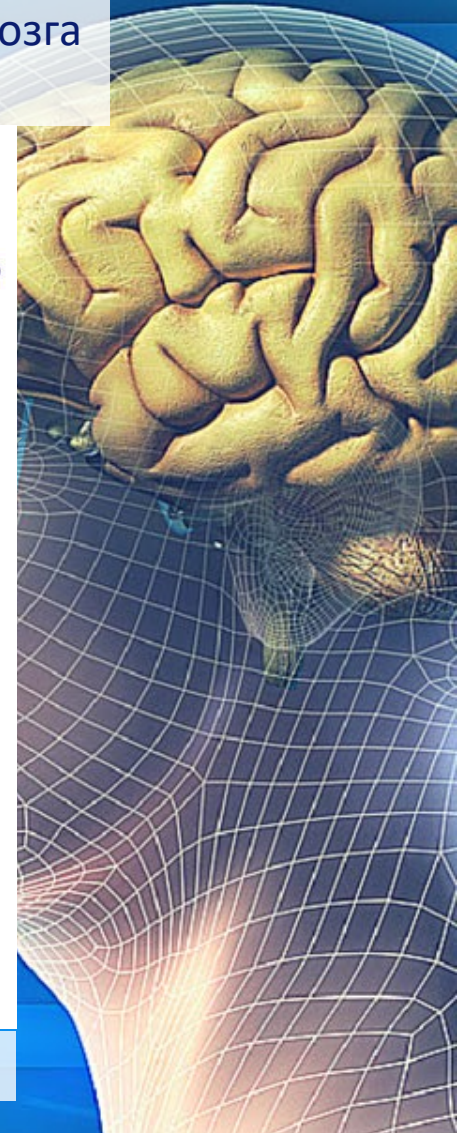
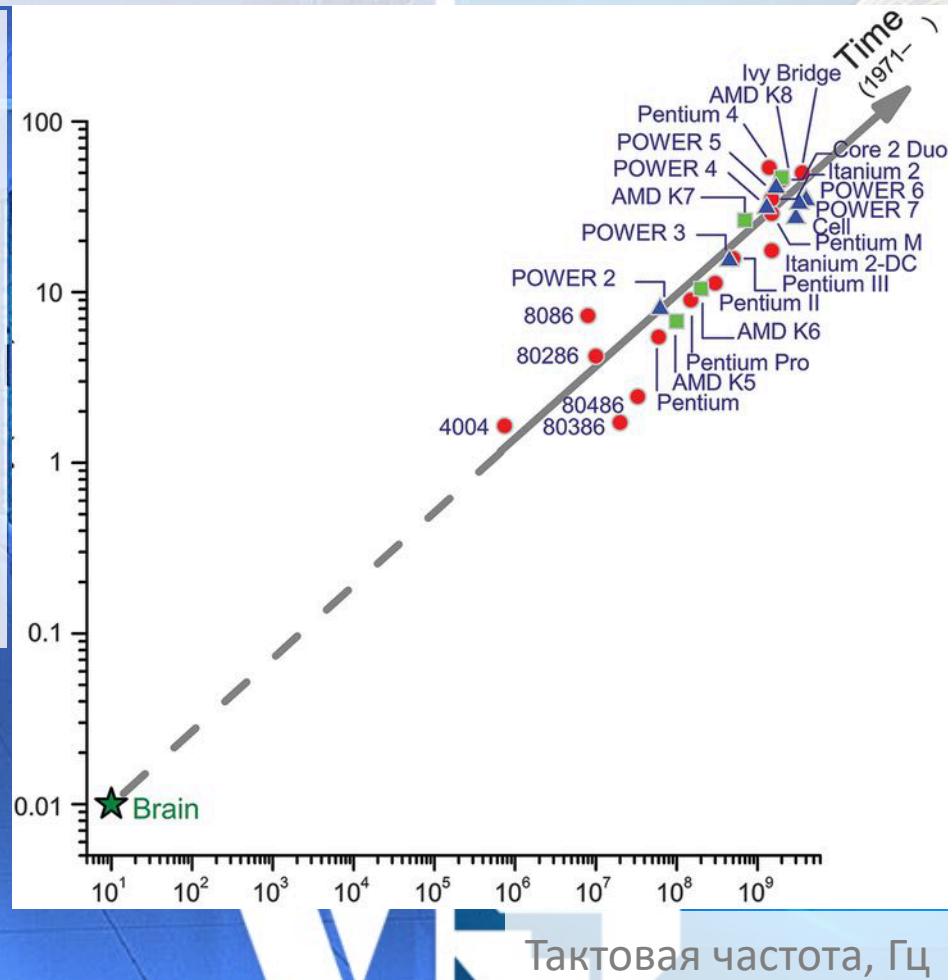
СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

Нейроморфная электроника на базе многоуровневой мемристорной логики

Мозг vs компьютер

Сравнение удельной мощности и тактовой частоты мозга с процессорами общего назначения

Плотность мощности, Вт/см²



Мозг vs компьютер

Мозг можно свести к искусственной нейронной сети (сети нейронов)

Для симуляции нейрональной активности требуется 10^{11} нейронов, у каждого по 10^4 синапсов, срабатывающих на частоте 10 Гц, требуемая производительность – 10^{16} операций в секунду.

Трудность в его реализации заключается в том, что на сегодняшний день не существует цифровой копии структуры мозга с клеточным разрешением.

Повышая уровень абстракции, работу мозга можно эмулировать посредством создания нейронных сетей соответствующей архитектуры с набором алгоритмов глубокого машинного обучения. Такие сети уже успешно применяются в системах компьютерного зрения, машинного перевода, распознавания речи.

VS

Мозг vs компьютер

Увеличение производительности современных суперкомпьютеров может использоваться для описания работы мозга, исходя из известных механизмов его функционирования, но при этом суперкомпьютер нельзя рассматривать в качестве эквивалента человеческого мозга вследствие различия его архитектуры и принципов обработки информации.

Предполагается:

- основная вычислительная единица мозга не нейрон, а простая нейросеть (или клик).
- простые нейросети уже существуют в мозге (их структура определяется генами).
- из кликов формируется более сложная нейросеть, при этом в процессе построения ее размерность постоянно меняется, начиная с 1D, размерность нейронных сетей может достигать до 11.
- нейронная сеть формируется на время решения задачи, после чего распадается.

Нейрокомпьютерные системы

Ограничение эффективности работы систем ИИ:

- замедление закона Мура об экспоненциальном развитии электроники с 2005 года [Chien-Ping Lu, NovuMind Inc. AI, Native supercomputing and the revival of Moore's law];
- узкое место архитектуры фон-Неймана, лимитирующее эффективное быстродействие современных процессоров при обработке больших объемов данных пропускной способностью между процессором и памятью.

Развитие нейрокомпьютерных систем:

- переход к массивным параллельным аналоговым вычислениям при организации искусственных нейронных сетей.

Использование классической архитектуры нейронных сетей с алгоритмами глубокого машинного обучения на существующей компьютерной архитектуре при распараллеливании потока входных данных приводит к увеличению времени обучения, проблемам маршрутизации данных, становится крайне неэффективным при работе с большими асинхронными потоками данных и требует большого энергопотребления



Нейрокомпьютерные системы

Спайковые нейронные сети прямого распространения рассматриваются в качестве наиболее физиологически реалистичной модели биологических нейронных сетей.

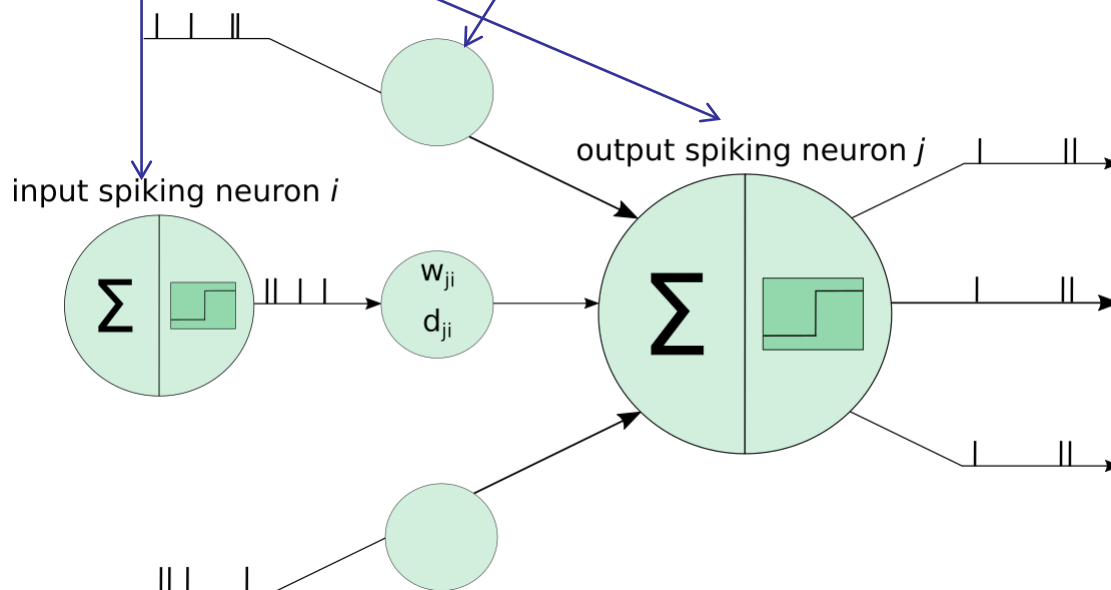
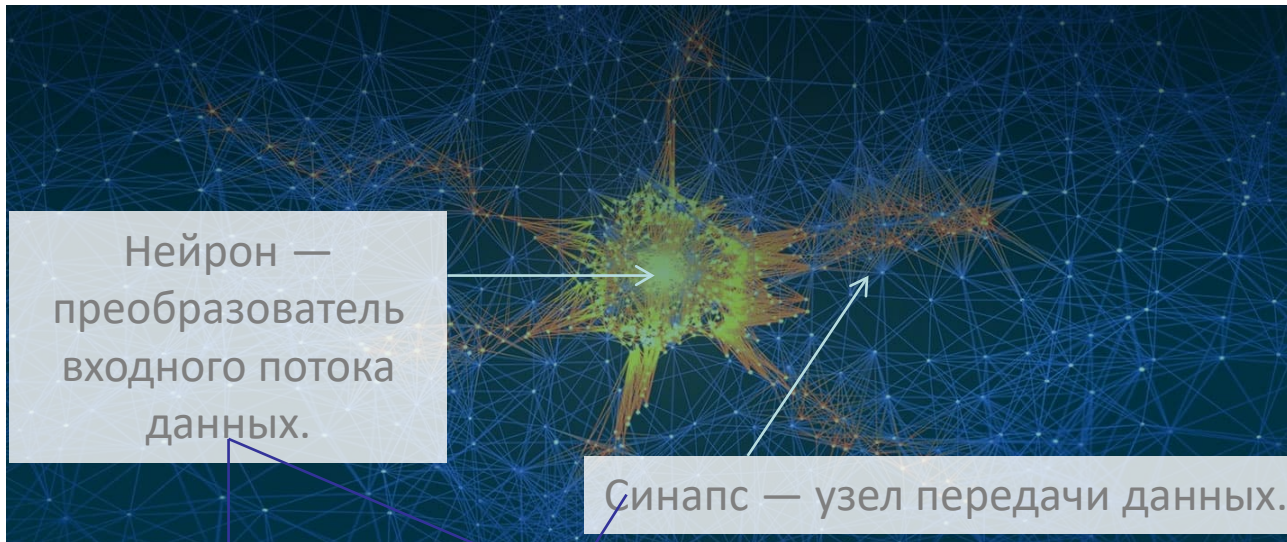
В таких сетях модели нейронов обмениваются между собой короткими импульсами, или спайками, распространение которых по сети определяется набором временных задержек.

Их наличие является ключевым моментом, поскольку позволяет обрабатывать информацию в соответствии с известными базовыми алгоритмами работы биологических нейронных сетей, а также имплементировать механизмы обучения и памяти.

Взаимосвязанность нейронов в спайковых нейронных сетях может быть смоделирована с использованием AER (address event representation) протокола *.

Спайк рассматривается как событие, которое несет информацию о времени своего возникновения. Поэтому они эффективно формируются в пакеты, содержащие адрес нейрона, генерирующего спайк, и могут использоваться для многоадресной передаче на разные нейроны.

Спайковые нейронные сети

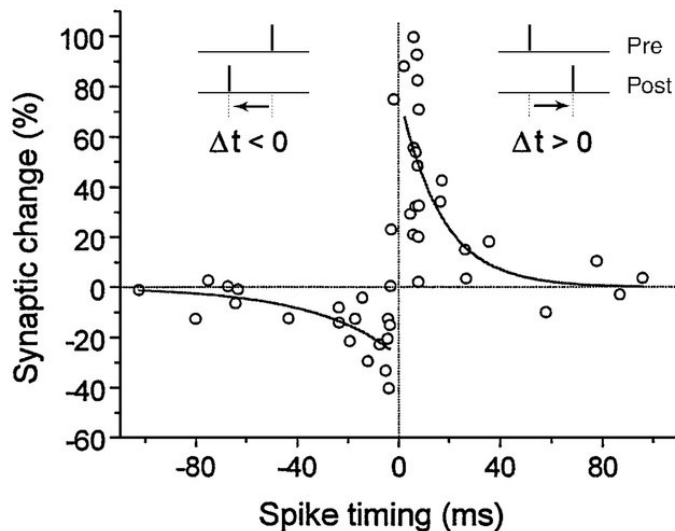


Самая приближенная модель ИНС с точки зрения физиологии.

Спайковые нейронные сети

Один из распространенных методов обучения — модель синаптической пластичности, зависящей от времени спайка (STDP).

Это возможность изменения силы синапса (веса коэффицента) в зависимости от времени генерации спайков.



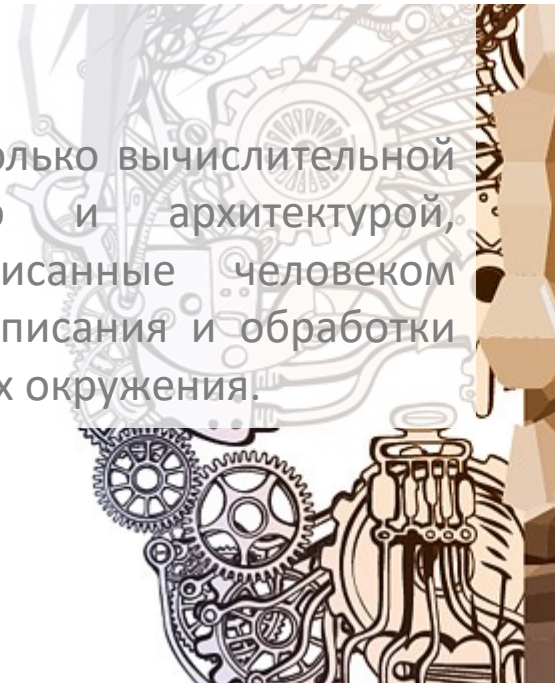
С. Дали. Постоянство памяти.

Синаптическая пластичность считается основным механизмом, который обеспечивает и обучение, и память сети. Такой механизм характерен для нервной системы человеческого мозга.

Нейрокомпьютерные системы

Программируемые машины

ограничены не только вычислительной мощностью, но и архитектурой, требующей написанные человеком алгоритмы для описания и обработки информации из их окружения.

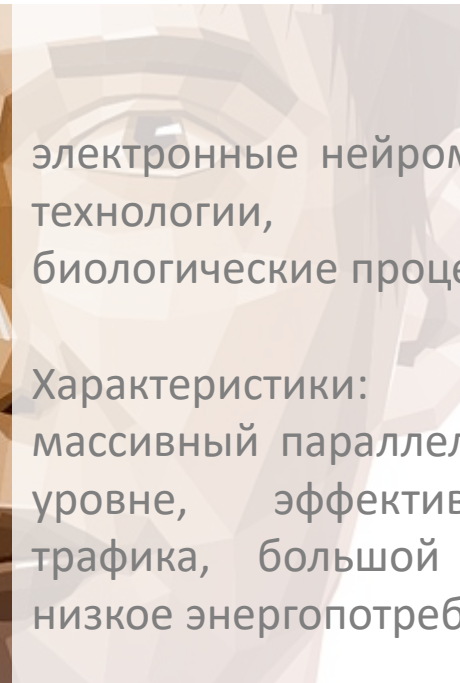


Нейрокомпьютерные системы

электронные нейроморфные машинные технологии, воспроизводящие биологические процессы.

Характеристики:

массивный параллелизм на аппаратном уровне, эффективная организация трафика, большой объем памяти и низкое энергопотребление.



Подходы к построению нейрокомпьютерных систем

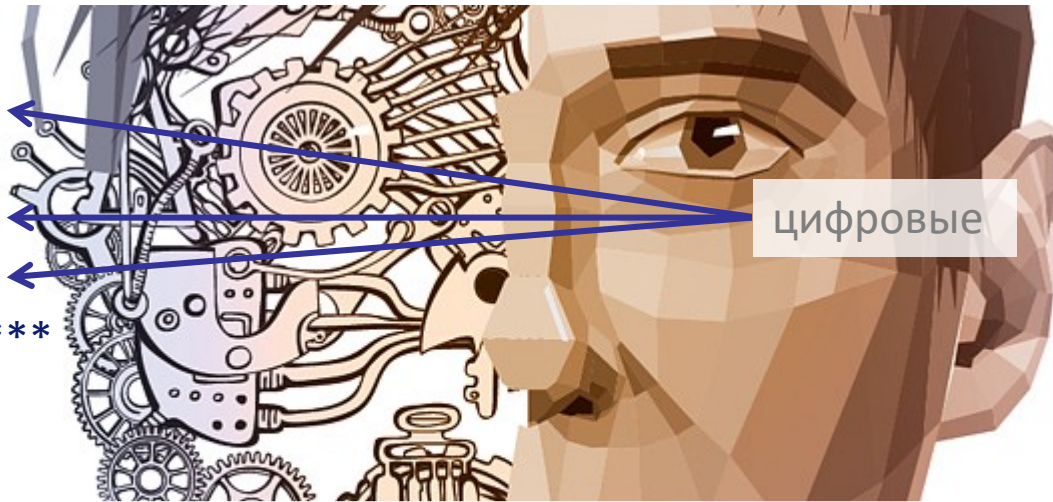
- Разработка нейроморфных модулей
- Использование цифро-аналоговых СБИС (СверхБольшая Интегральная Сема)
- Организация неспециализированных вычислений на графических процессорах
- Использование FPGA – ускорителей (ускорителей на основе программируемых логических интегральных схем)
- Использование нанотехнологий для создания интеллектуальных материалов.
- Комбинация вышеперечисленных подходов.



Нейрокомпьютерные системы

Широко известными прототипами нейроморфных процессоров, работающими на существующей компонентной электронной базе, являются:

- Neurogrid *
- TrueNorth **
- DYNAPs ***
- SpiNNaker ****
- Loihi *****
- Braindrop *****



Глобально асинхронная локально синхронная архитектура - модель вычислений, появившаяся в 1980-х годах, которая позволяет проектировать компьютерные системы, состоящие из нескольких синхронных островов (с использованием синхронного программирования для каждого такого острова), взаимодействующих с другими островами с использованием асинхронной связи.

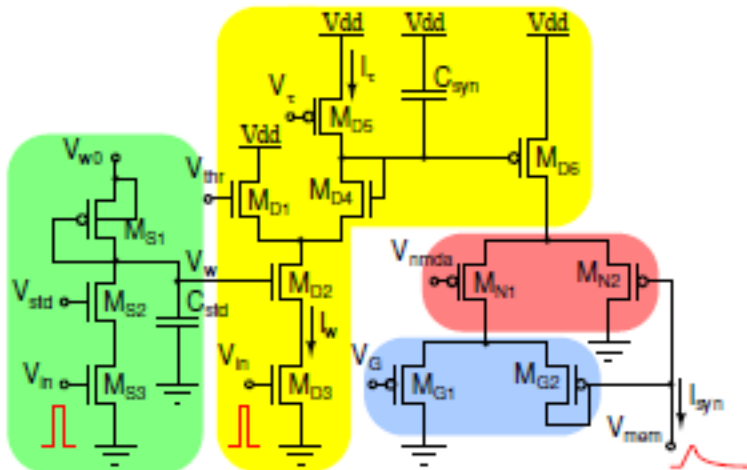
Нейроморфные процессоры

Цифровые

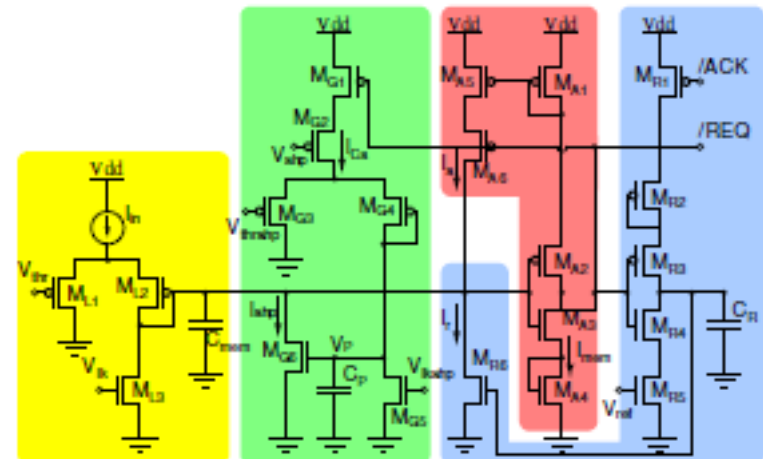
- **Intel Loihi** (Полностью асинхронная нейроморфная многоядерная сетка, на основе технологии 14 нм Intel, 130 000 нейронов и 130 миллионов синапсов)
- **TrueNorth** (NS16e-4 — самый крупный сегодня нейросинаптический компьютер: 64 млн нейронов и 16 млрд синапсов; выполняет порядка 1011 синаптических операций за секунду в расчете на ватт; общая потребляемая мощность составляет 70 Вт, из которых 15 Вт расходуются всеми 64 процессорами TrueNorth.)

Аналоговые

Complete DPI synapse circuit



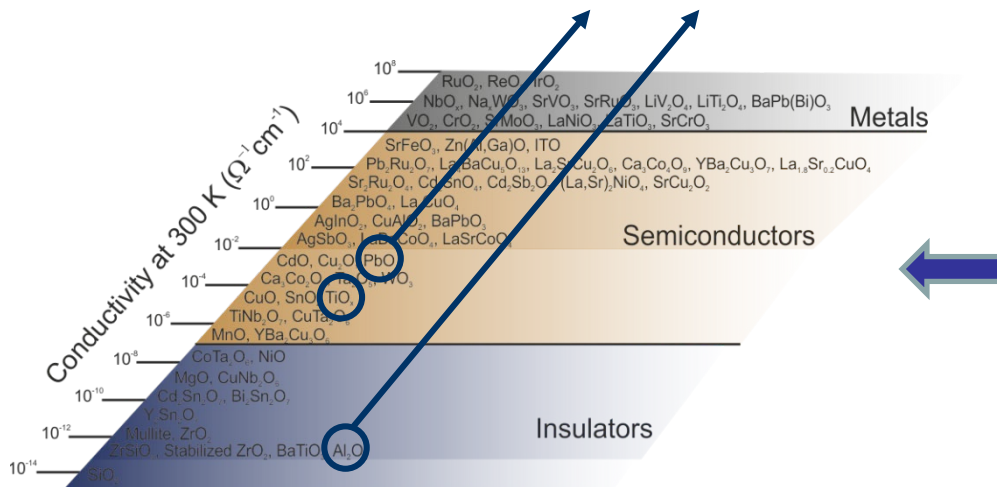
Adaptive exponential I&F neuron



Интеллектуальные материалы для нейроморфных процессоров

PbO, TiO₂, Al₂O₃, TiO₂/Al₂O₃

(тонкопленочные, толщинами от единиц до десятков нанометров, металлооксидные структуры)



- adapted from P. Edwards, V. Kuznetsov, D. Slocombe and R. Vijayaraghavan, *Comprehensive Inorganic Chemistry II*, 2013.

Использование новой компонентной базы на основе интеллектуальных материалов обладает рядом преимуществ, таких как:

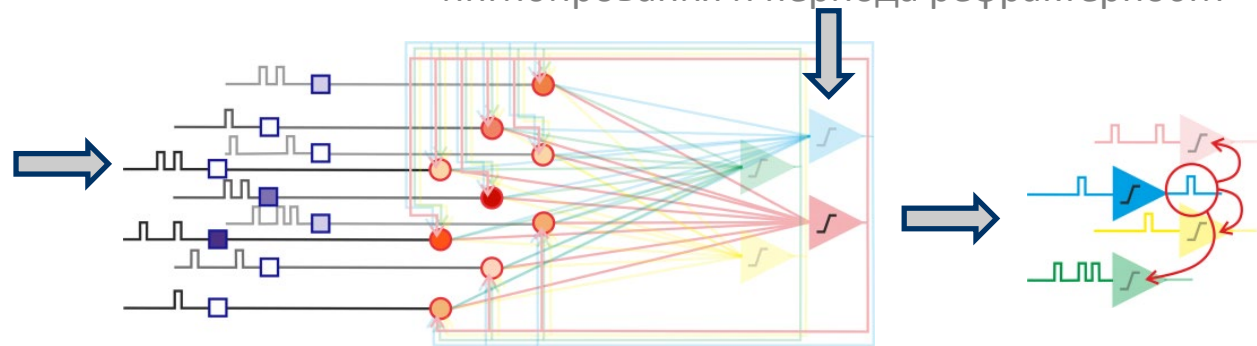
- более высокие скорости переключения между высокоомным и низкоомным состояниями,
- большая устойчивость к циклической деградации, меньшее энергопотребление,
- возможность создания многоуровневых состояний.

Аналоговые нейроморфные процессоры

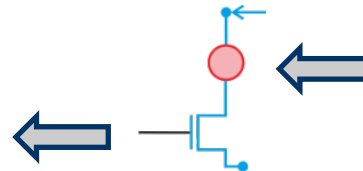
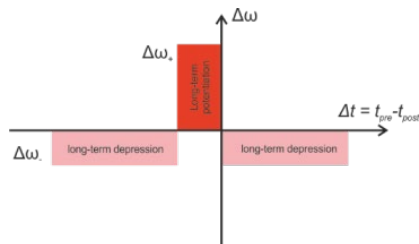
пластичный нейрон

пороговый интегратор с утечкой и адаптивным порогом, функциями латерального ингибирования и периода рефрактерности

Входной поток
асинхронных данных
с сенсоров.



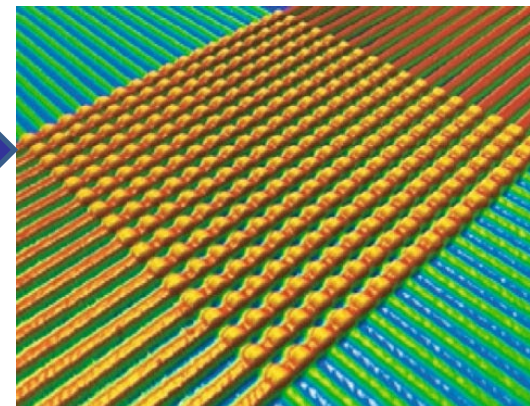
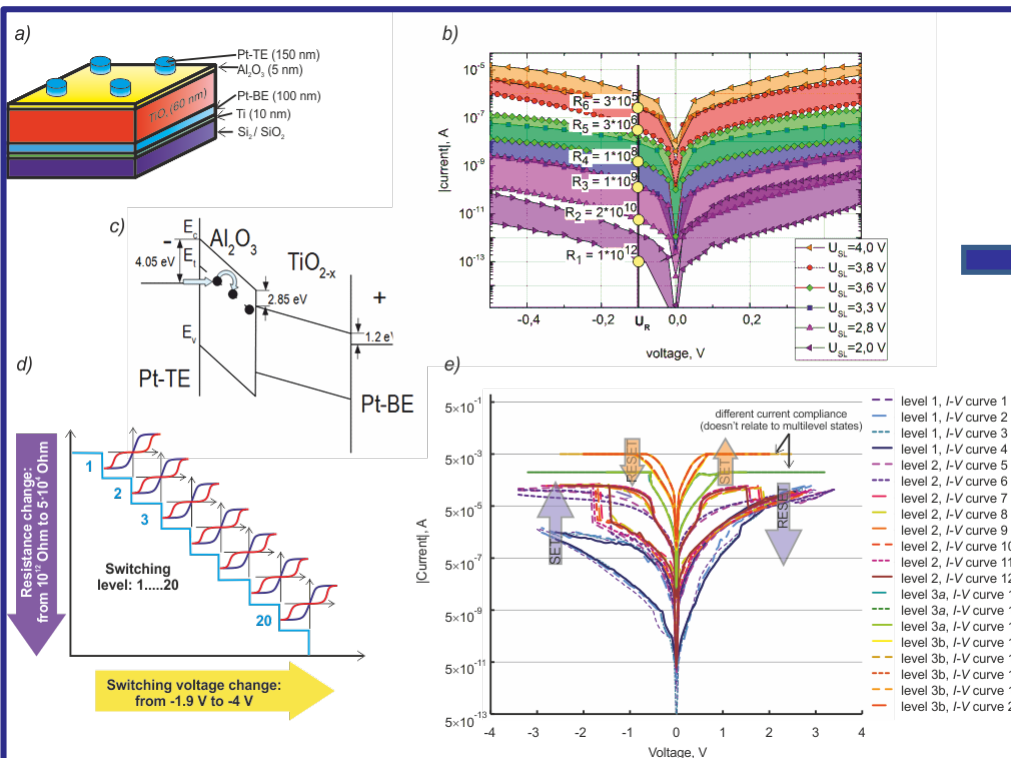
Хеббиановские алгоритмы обучения
(Хеббиановская пластичность)



синапс на базе
мемристовой
тонкопленочной
структуры

Мемристивные материалы с многоуровневым переключением сопротивления

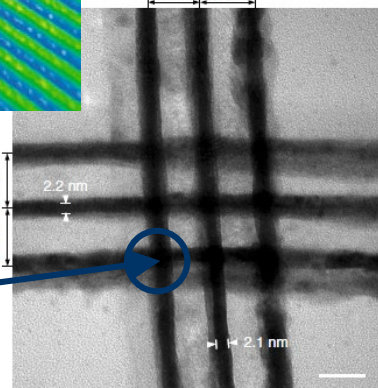
Использование в структуре «crossbar»*



Top nanofin array
11.5 nm 12.0 nm

Bottom nanofin array
11.9 nm 11.3 nm

2 nm x 2 nm



Мировые научные лидеры



СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

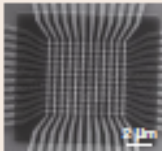
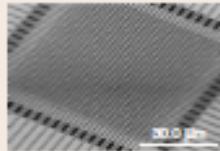
HP + MIT

Stanford +
Beijing

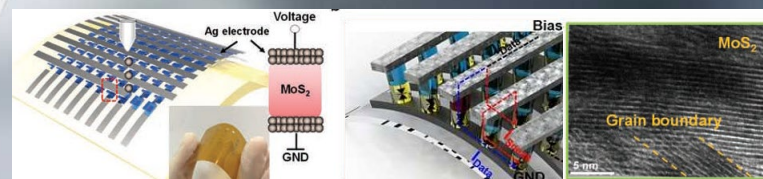
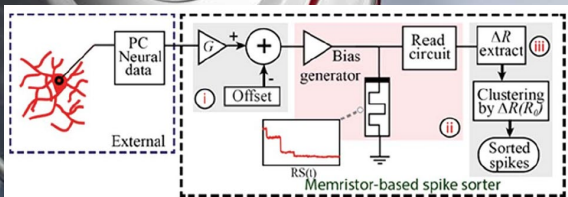
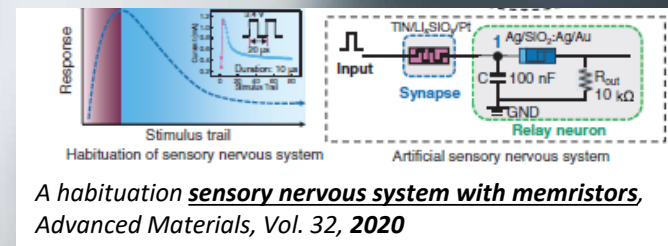
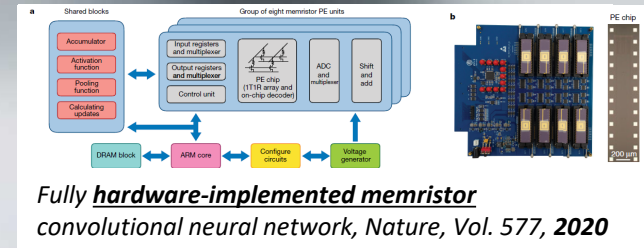
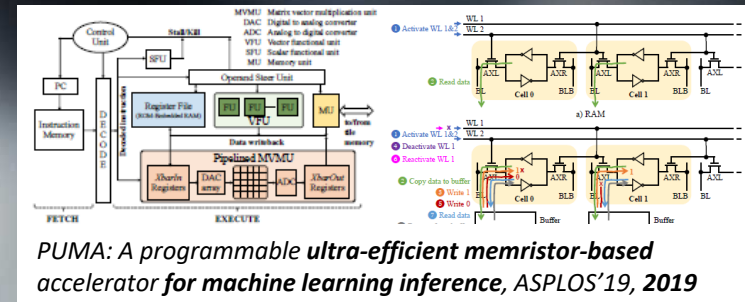
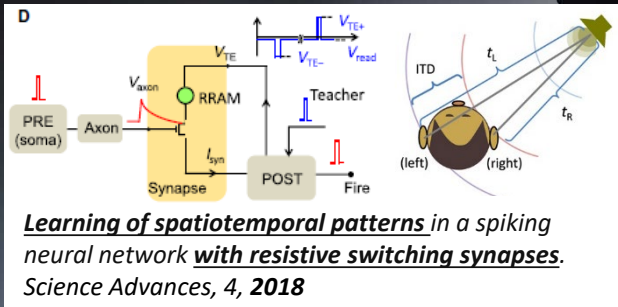
CA +
Japan

Michigan
(USA)

Southampton +
Graz +
Heidelberg

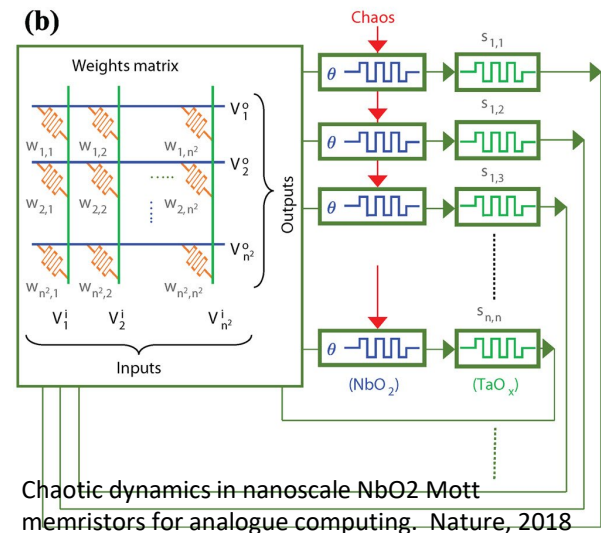
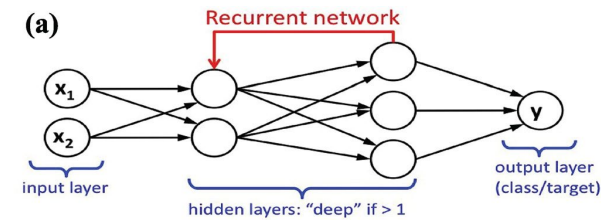
	Active arrays (1T1R)		Passive arrays		
Memristor geometry	Ta/HfO ₂ /Pt or Pd	TiN/HfAlO ₂ /TaO ₂ /TiN	Pt/Al ₂ O ₃ /TiO _{2-x} /Ti/Pt	W/WO ₂ /Pd/Au	Pt/TiO ₂ /Pt
Memristor size	4 × 4 μm ²	NA	200 × 200 nm ²	500 × 500 nm ²	NA
Transistor node	2 μm	1.2 μm	No transistors		
Max. array size	128 × 64	128 × 8	20 × 20	32 × 32	4 (wired) + 4 (software)
Array Image	 				NA
Multilevel mechanism	Composition modulation of a Ta-rich conduction channel	Modulation of the local concentration of oxygen vacancies	Reversible modulation of the concentration profile of oxygen vacancies	Modulation of the area ratio between a Schottky and a tunnelling junction	NA
Device retention	>>10 years at 85 °C	NA	>50,000 seconds	Tunable	2.5 hours
Conductance range	100–900 μS	10–60 μS	12–142 μS	0–5 μS	145–175 μS
I–V linearity	Linear	Nonlinear	Nonlinear	Nonlinear	NA
Synapse precision	0.75% error	5% or 35% error	7.4% error	5.6% error*	NA
Weight update scheme	Set/reset/gate amplitudes	Identical pulses	Identical pulses	Identical pulses	Set/reset amplitudes
Symmetry in weight update	Symmetric	Asymmetric	Asymmetric	Asymmetric	Asymmetric
Peripheral circuitry	Off-chip MCU with DAC/ADC, TIA	SPA + switching matrix	SPA + switching matrix + PCB hardware neuron	Off-chip MCU with DAC/ADC, TIA	Off-chip MCU with DAC/ADC, TIA
Application demonstrations	Image compression/ filtering; MNIST classification; time-series regression; gait recognition; reinforcement learning	Face classification	3 × 3 or 4 × 4 bitmap classification	Sparse coding; reservoir computing	Unsupervised learning

Нейроморфные мемристорные модули



Архитектура нейроморфных мемристормых модулей

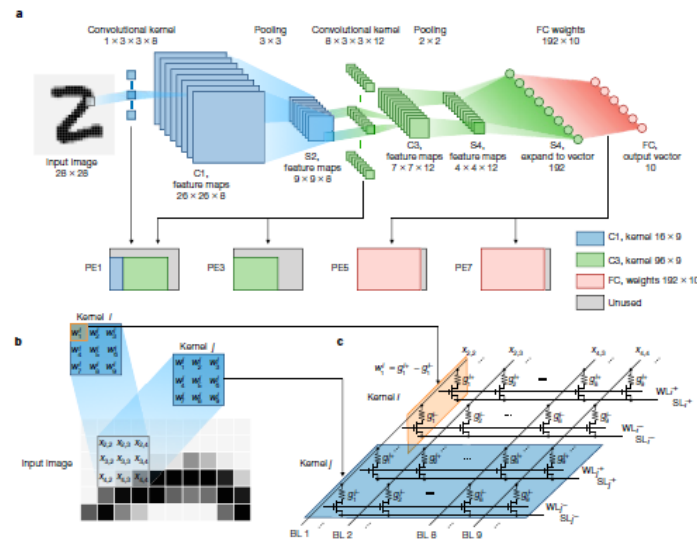
Рекуррентные сети Хопфилда на мемристормых кроссбарах



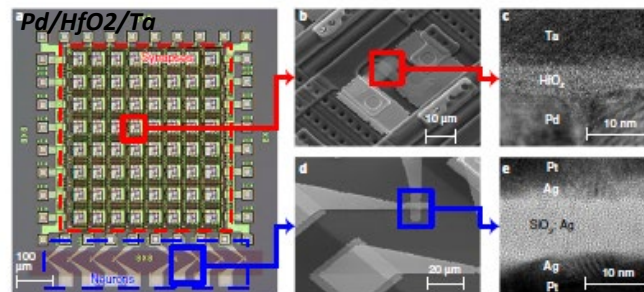
Chaotic dynamics in nanoscale NbO₂ Mott memristors for analogue computing. Nature, 2018

- University of Massachusetts
- Loughborough University
- Hewlett Packard Labs
- Air Force Research Lab, USA
- Tsinghua University, China

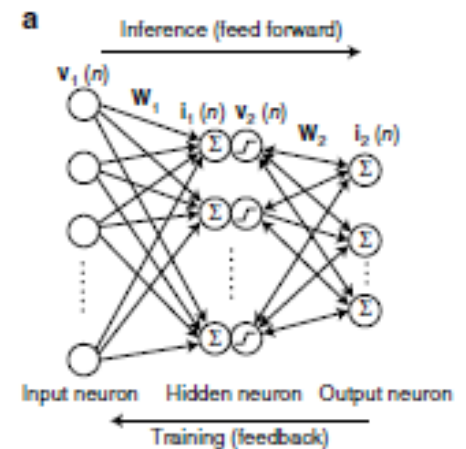
Сверточные сети на мемристормых кроссбарах



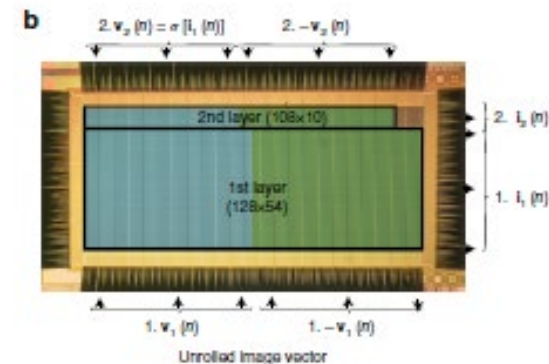
Fully memristive neural networks for pattern classification with unsupervised learning. Nature Electronics, 2018



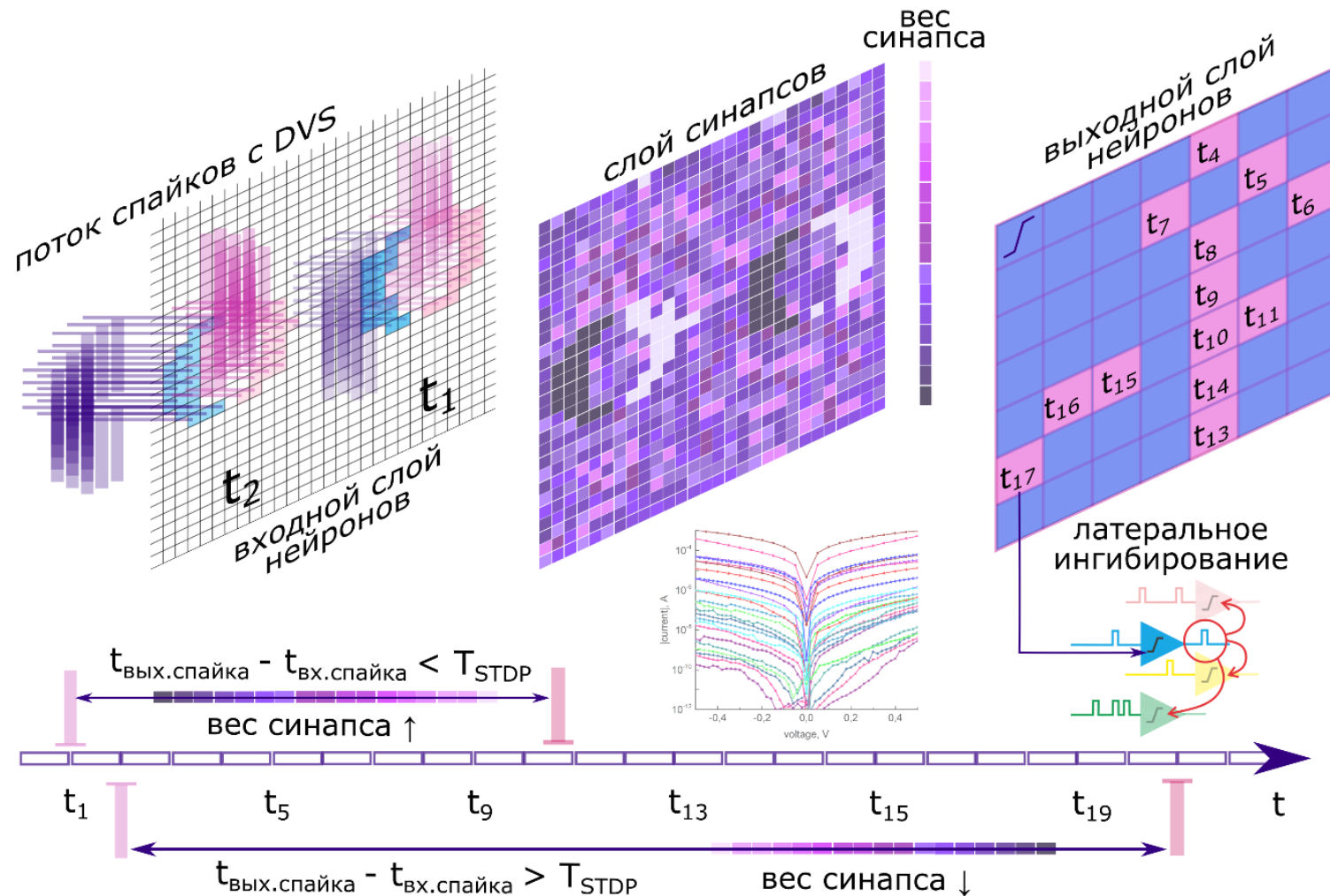
Многослойные перцептроны на мемристивных кроссбарах



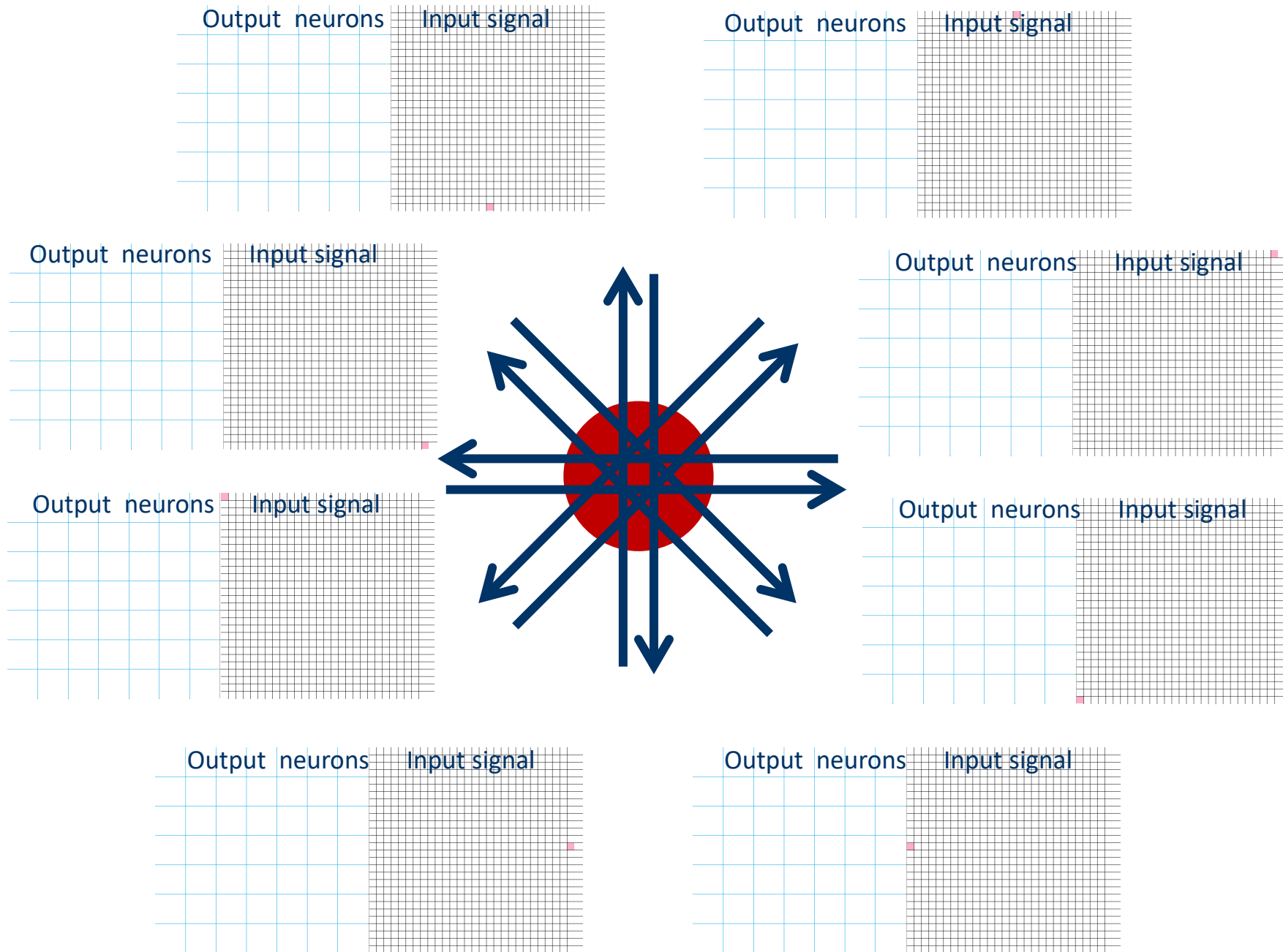
Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. Nature Communications, 2020



Спайковые нейронные сети прямого распространения с обучением без учителя на базе мультитбитных мемристоров.



Моделирование работы нейронной сети для распознавания траекторий движения сферы в поле зрения динамического видеодатчика

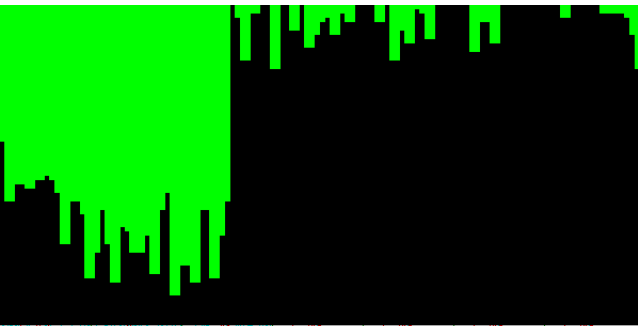


Моделирование работы нейронной сети для распознавания звуковых паттернов



Frequency spectrum

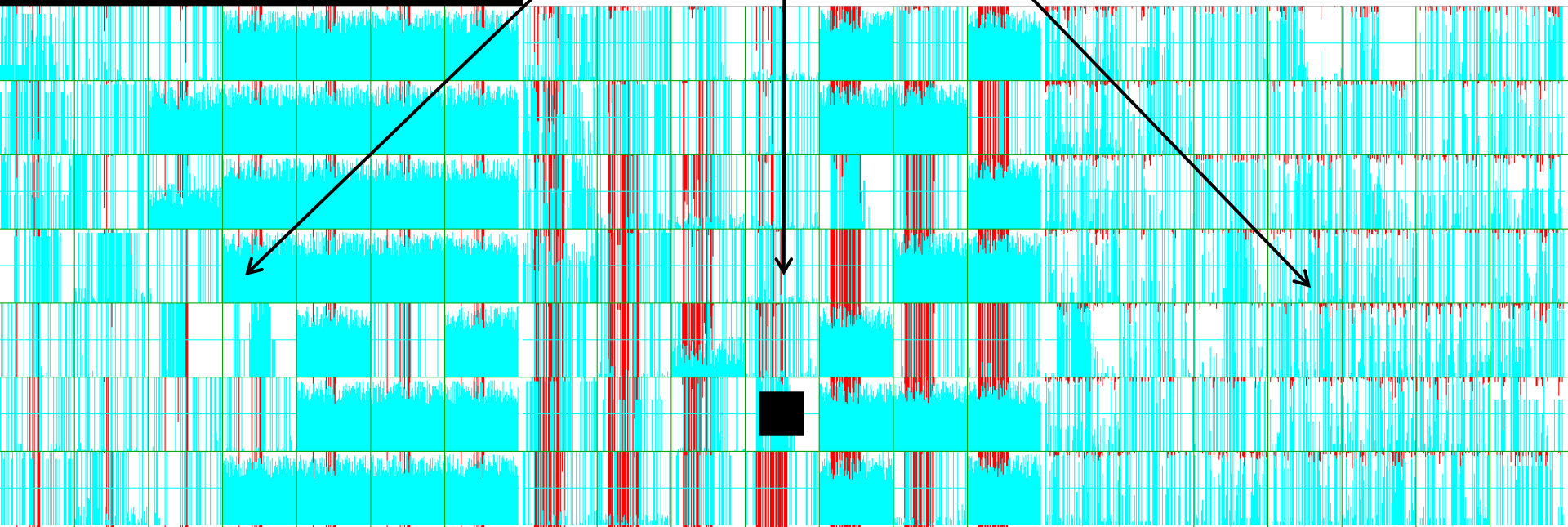
for signal window 32 msec,
8 kHz sampling rate



SNN for audio pattern recognition:

- cough;
- baby cry;
- glass;
- scream;
- siren

Output layer of neurons



Threshold for input signal: 3/4

1/2

1/3

Преимущества разработанной архитектуры спайковой нейронной сети

- ✓ выбранная **архитектура** спайковой нейронной сети является **универсальной**, т.е. может использоваться для решения различного класса задач.
- ✓ **асинхронный режим** работы обеспечивает существенное **снижение энергопотребления** нейроморфного модуля.

Рабочие параметры нейронной сети,
определенные из результатов компьютерного моделирования

Тип входных данных	Параметр спайковой нейронной сети				
	I_{TH} , тыс.	T_{LTP} , мс	T_{REF} , мс	T_{INH} , мс	τ_{leak} , мс
Определение траектории движения сферы	40	2	10	1,5	128
Распознавание звукового паттерна	2500	128	192	44,8	179,2

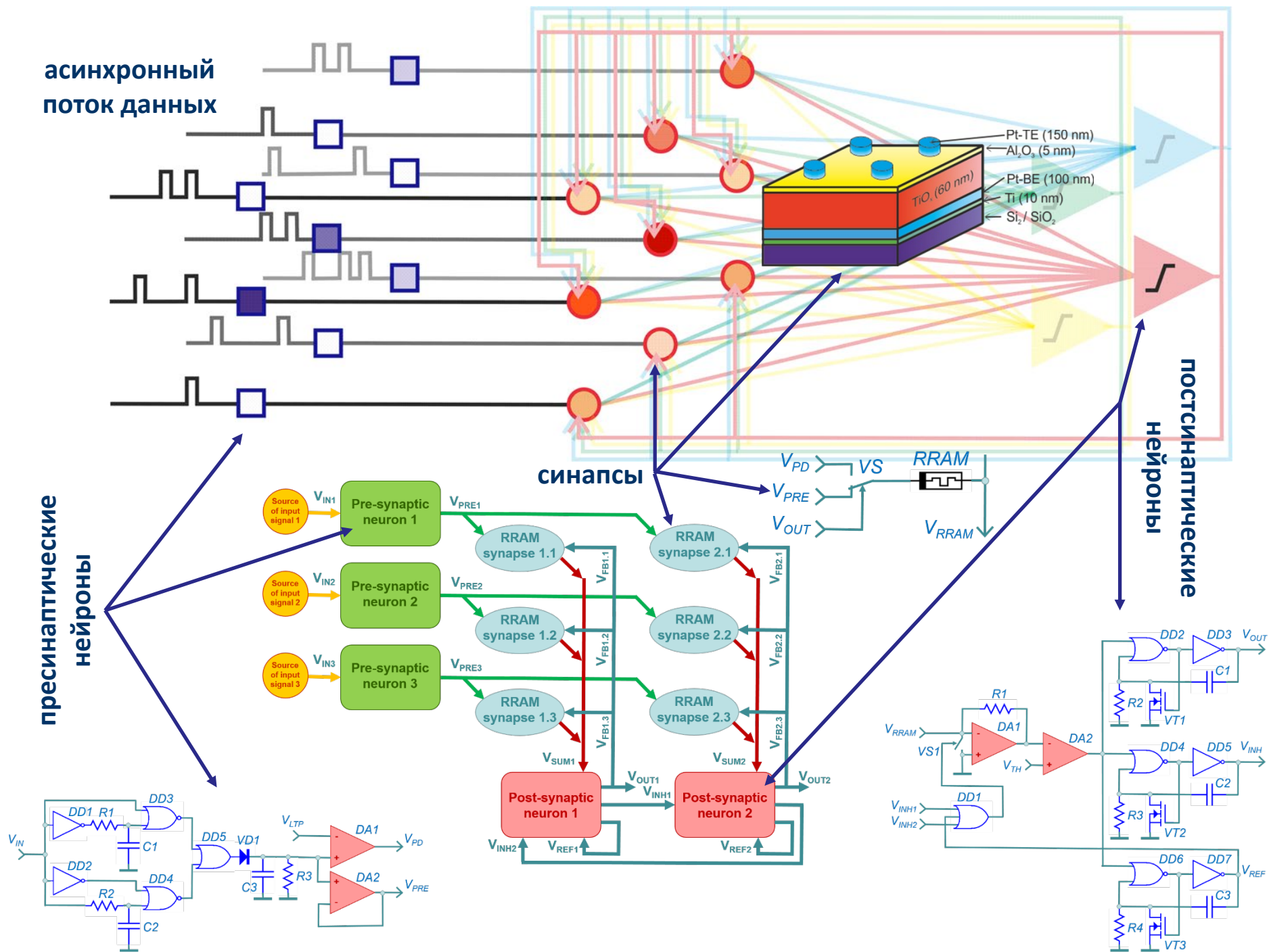
- ✓ обеспечение опции мультимодальности нейроморфного модуля на базе мемристорной логики предусматривает разработку решений узлов для **адаптивной самонастройки** основных рабочих параметров сети.

асинхронный
поток данных

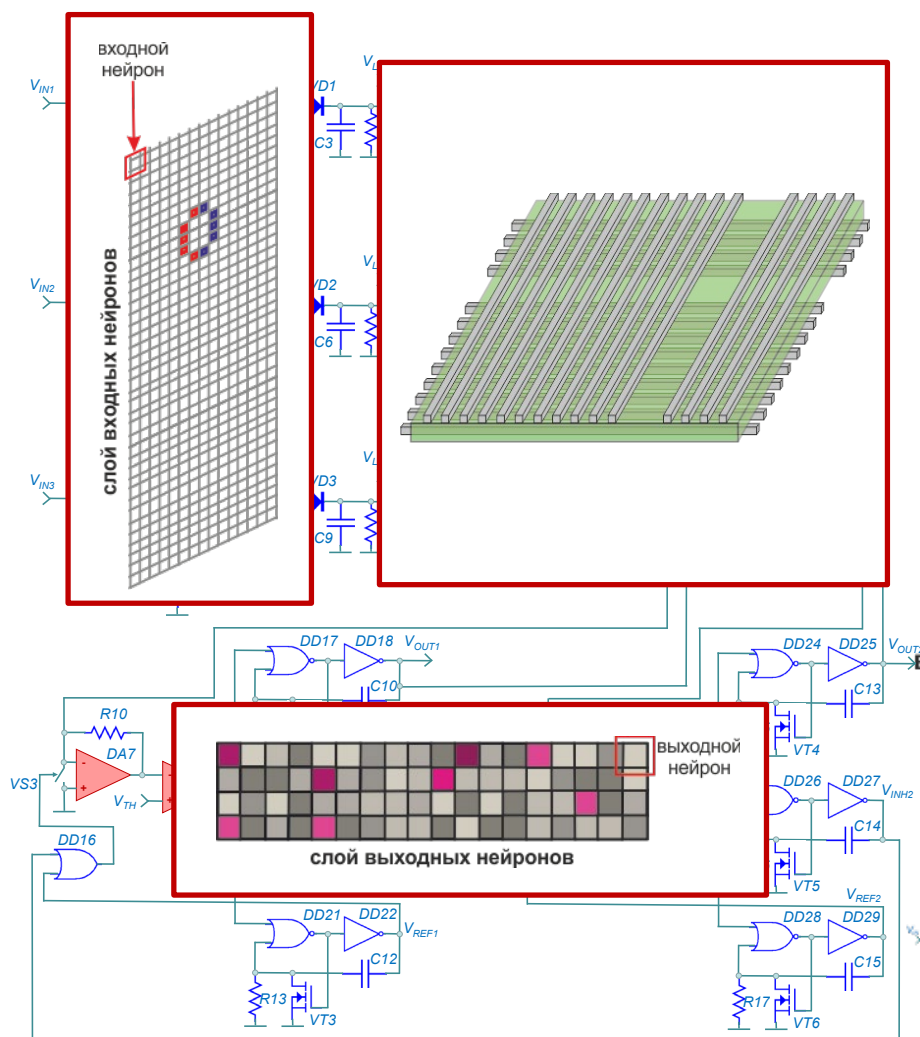
пресинаптические
нейроны

синапсы

постсинаптические
нейроны



Архитектура перцептрона на мемристивном кроссбаре



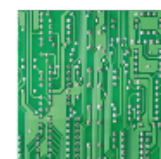
Преимущества:

- Сеть прямого распространения (отсутствует необходимость анализа хаотических состояний сети, проще схемотехническая реализация);
- Полносвязная сеть (выигрыш в энергопотреблении – работают только те связи, по которым распространяется сигнал);
- Большой динамический диапазон при решении задач ИИ, возможность решения разного рода задач, универсальность (каждому входному паттерну соответствует своя конфигурация).

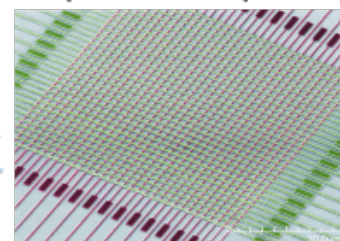
интегральная схема
входного слоя нейронов



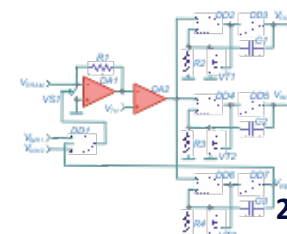
интегральная схема
выходного слоя нейронов



мемристивный кроссбар



<https://news.engin.umich.edu/features>



Много-уровневые мемристоры

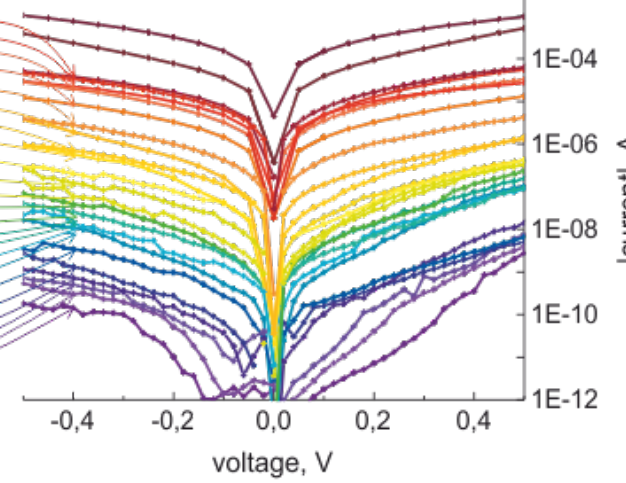
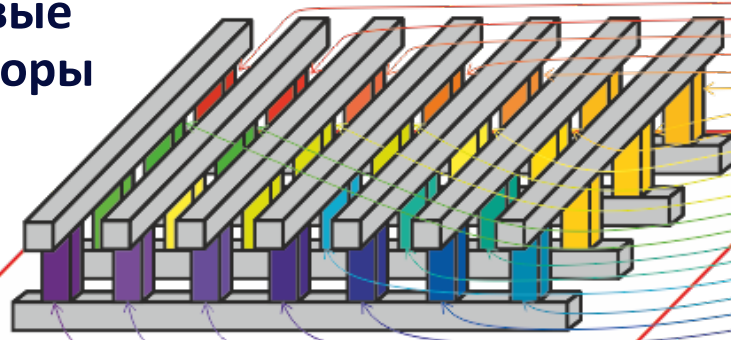
распределение логико-информационных уровней мемристивных структур в устройстве кроссбара

вес синаптической связи

аналоговая многоуровневая перестройка сопротивления мемристивной гетероструктуры

входной нейрон

слой входных нейронов



N2

N1

M1 x M2

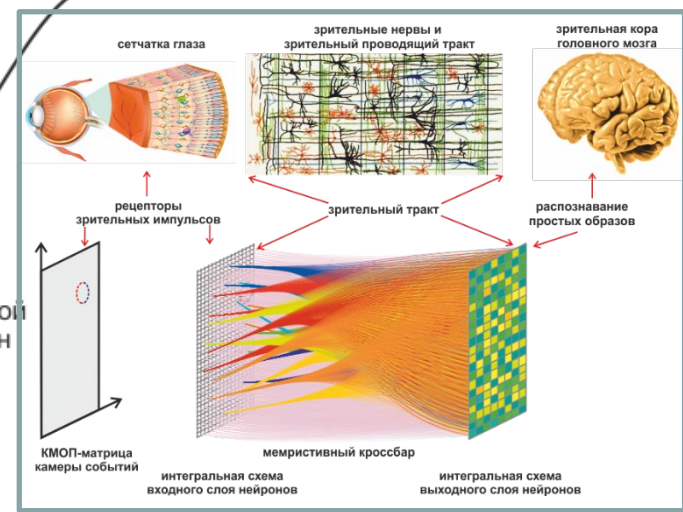
N1 x N2

порог срабатывания выходного нейрона

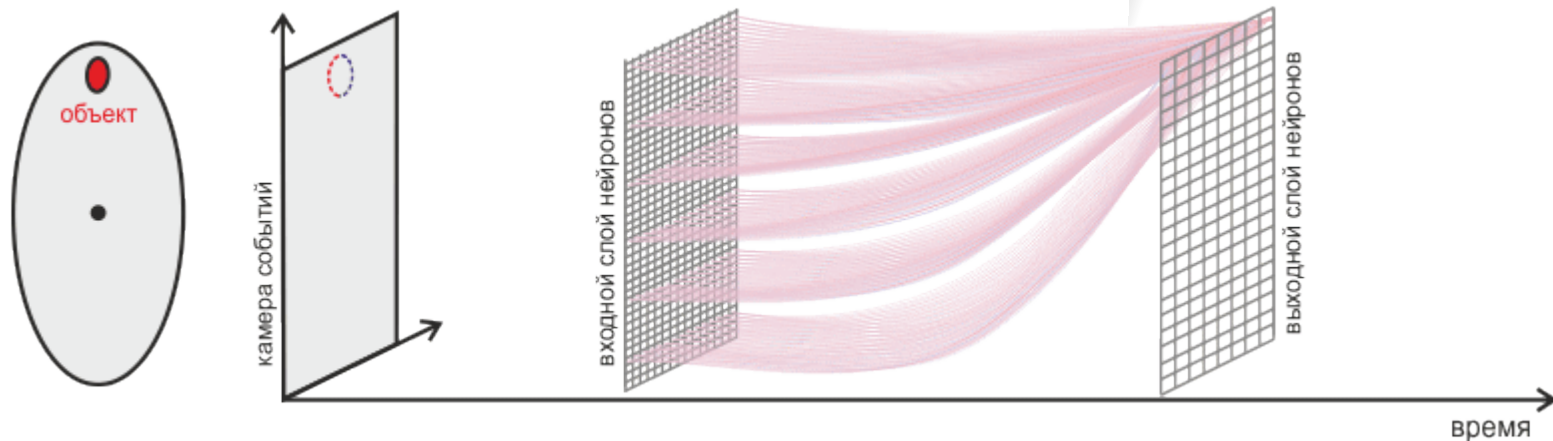
M1

выходной нейрон M2

слой выходных нейронов



Архитектура перцептрона на мемристивном кроссбаре для задач распознавания направления движения

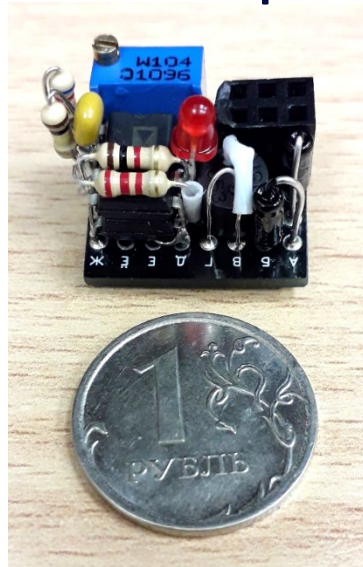


Спайковые нейронные сети прямого распространения:

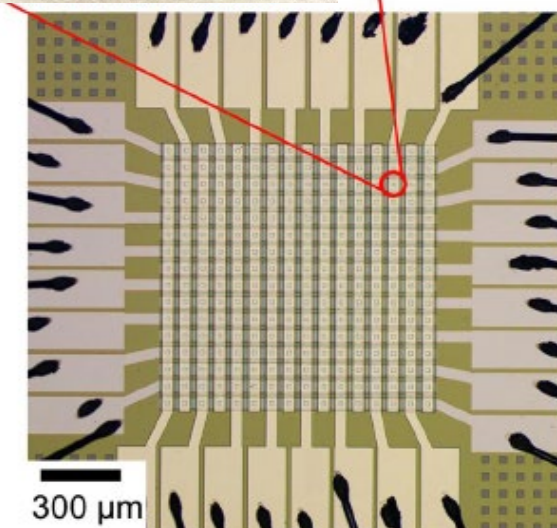
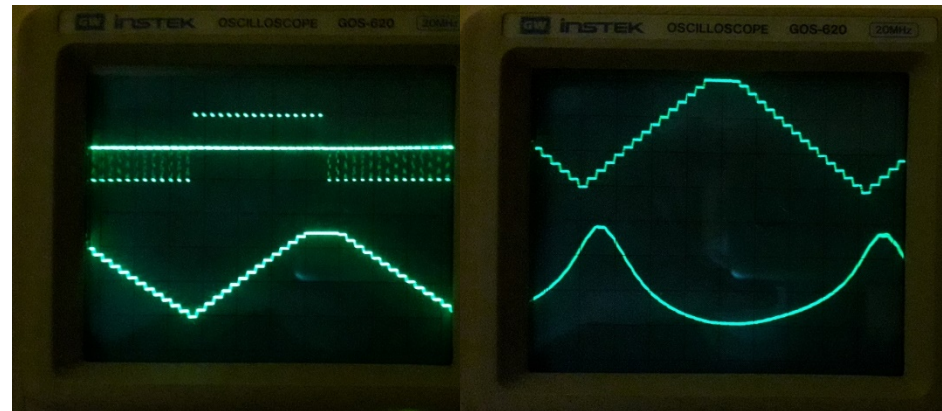
- Рассматриваются в качестве наиболее физиологически реалистичной модели биологических нейронных сетей.
- В таких сетях модели нейронов обмениваются между собой короткими импульсами, или спайками, распространение которых по сети определяется набором временных задержек. Их наличие является ключевым моментом, поскольку позволяет обрабатывать информацию в соответствии с известными базовыми алгоритмами работы биологических нейронных сетей, а также имплементировать механизмы обучения и памяти.
- Для обучения спайковых нейронных сетей используются хеббоподобные правила, определяющие усиление или ослабление синаптической связи между нейронами, и являющиеся, по сути самоорганизующимися алгоритмами, реализующими обучение без учителя.

Макетирование функциональных блоков нейроморфных модулей

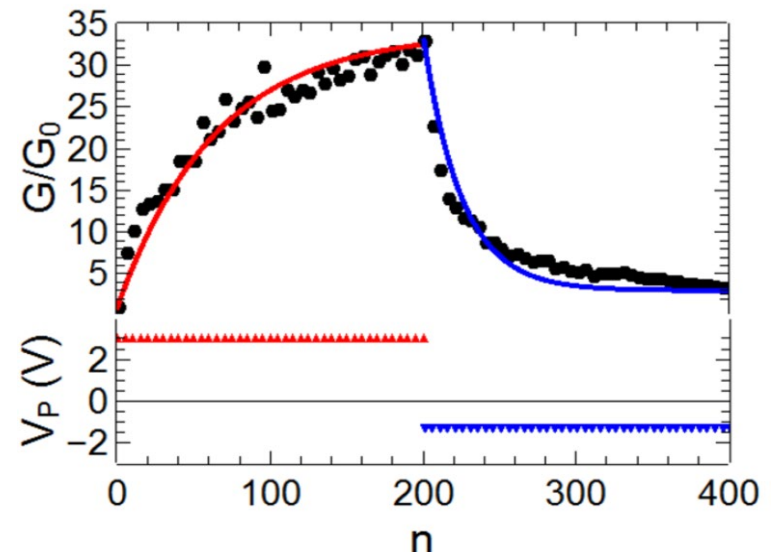
Функциональный
аналог мемристора



Управление сопротивлением
функционального аналога мемристора



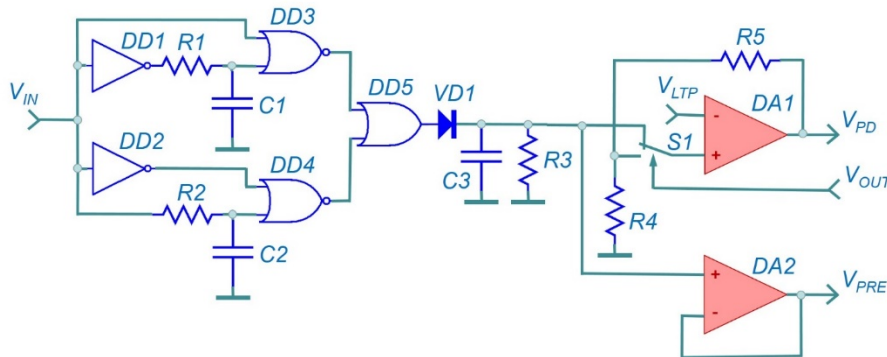
Мемристорная cross-bar матрица



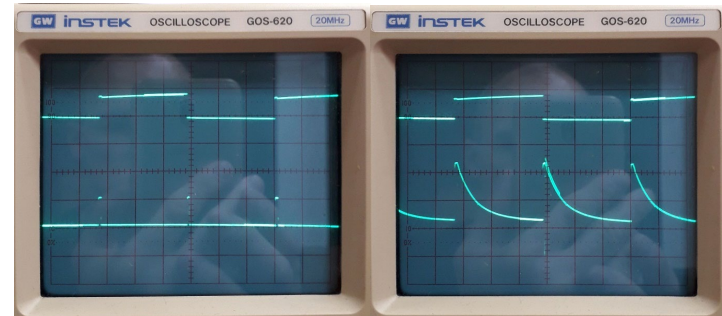
Управление сопротивлением мемристора

Макетирование функциональных блоков нейроморфных модулей

Схема пре-нейрона



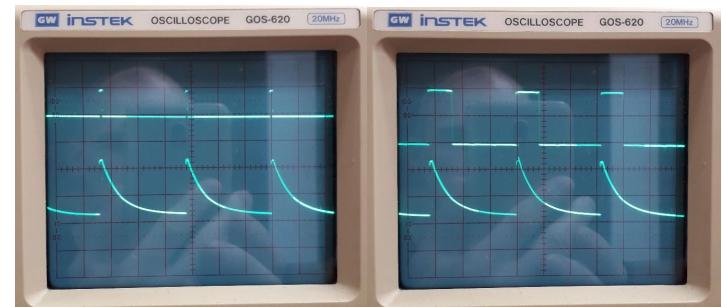
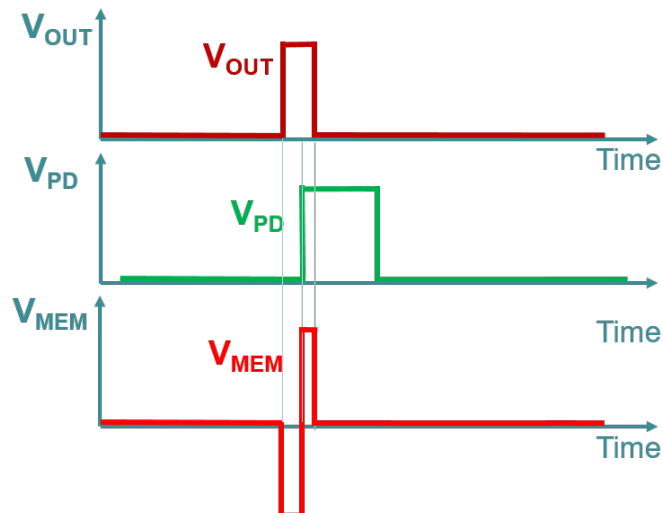
Осциллограммы макета пре-нейрона



а

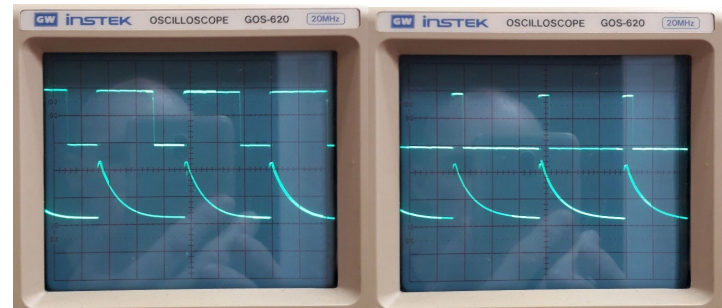
б

Формирование разнополярных
импульсов потенции/депрессии



в

г



д

е



Спасибо за внимание!