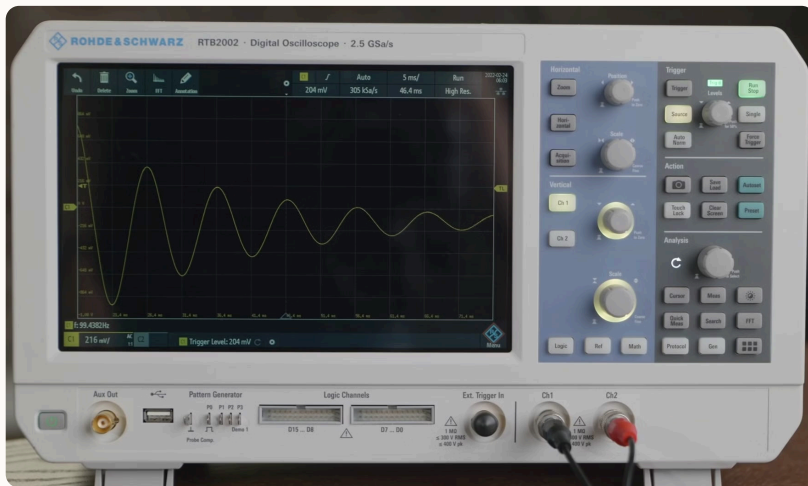


Использование аналоговых систем в сфере искусственного интеллекта

Балин Артем Алексеевич, Р3212

Рассматриваемые проблемы

- Как можно применять аналоговые системы на практике при работе с ИИ
- Преимущества и недостатки аналоговых систем
- Примеры их реализации

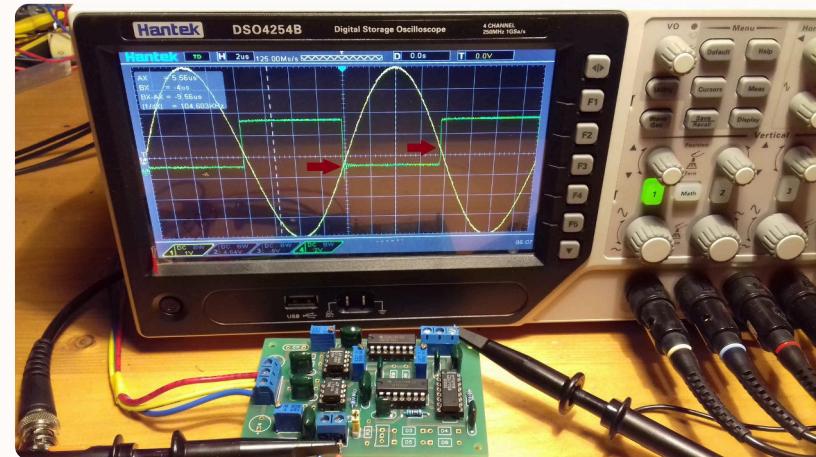


Повторение

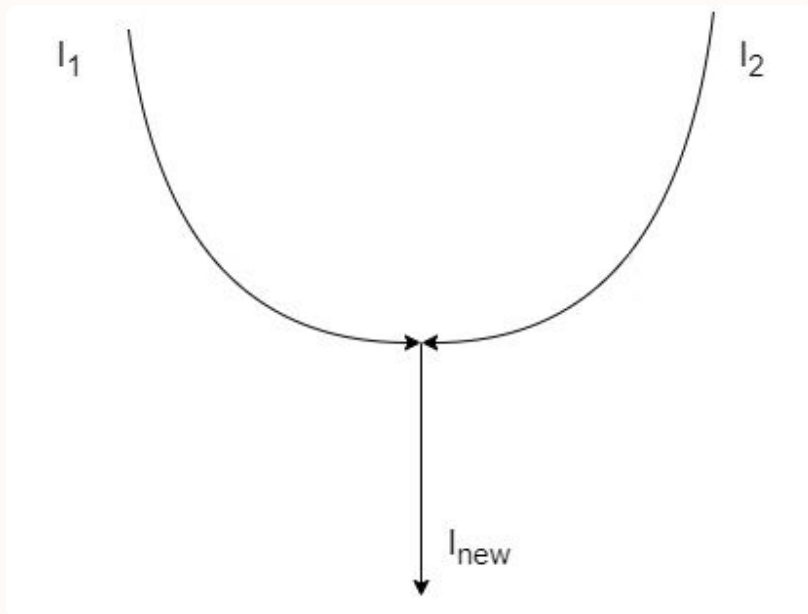
Аналоговые системы - системы, в которых в качестве сигналов выступают непрерывные величины

Провода и осциллограф - плохо?

- Нельзя задать точные входные параметры
- Невозможно повторно использовать
- Погрешность вычисления

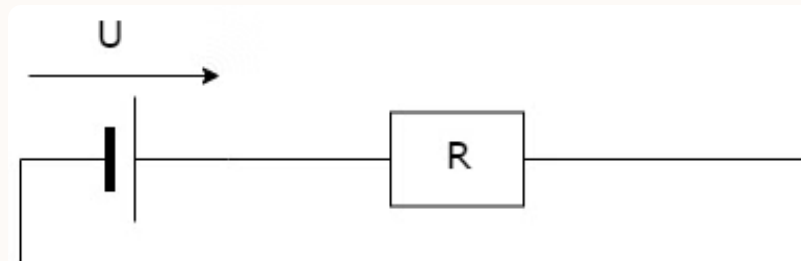


Как реализовать арифметику?



Сложение

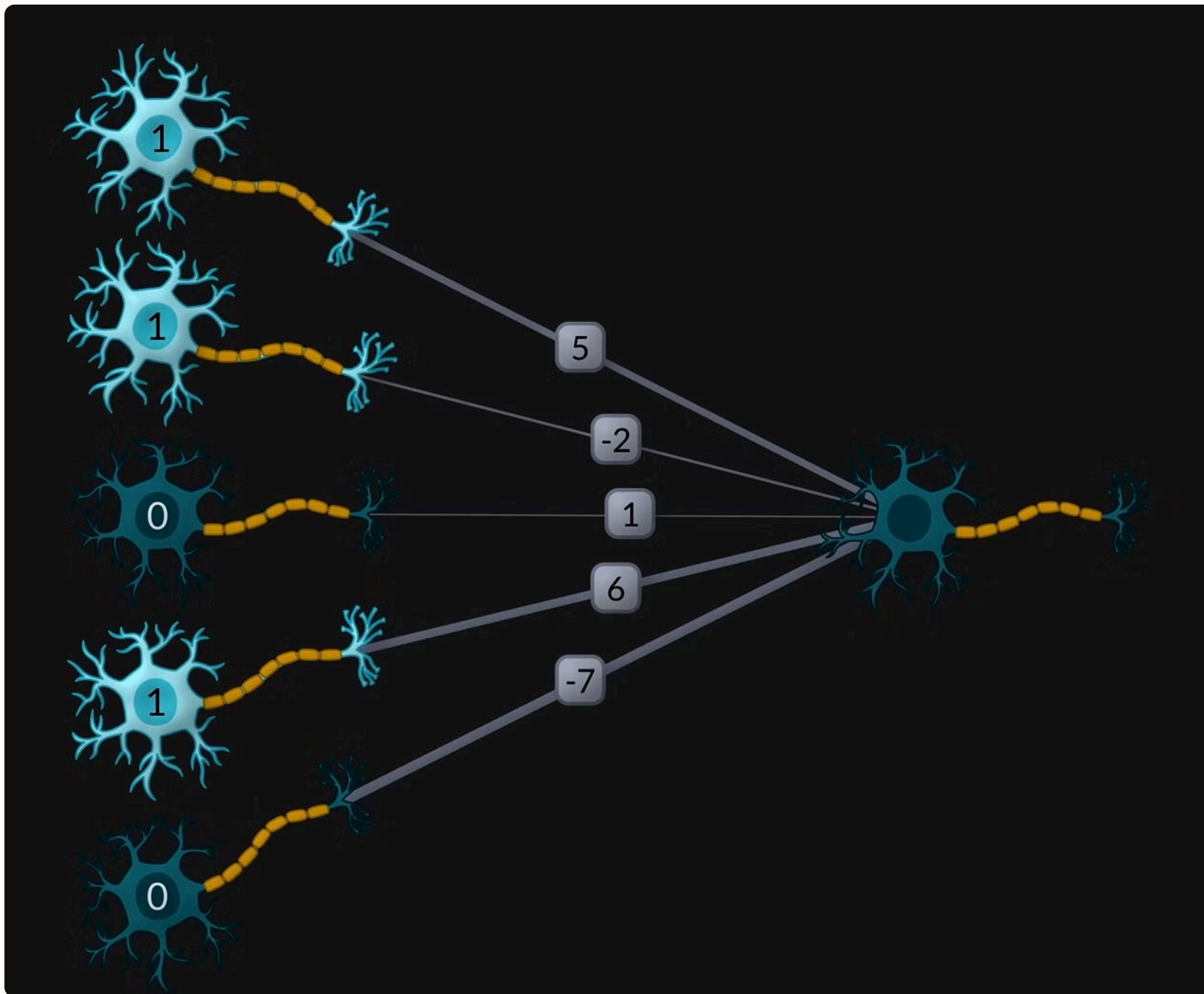
$$I_{new} = I_1 + I_2$$



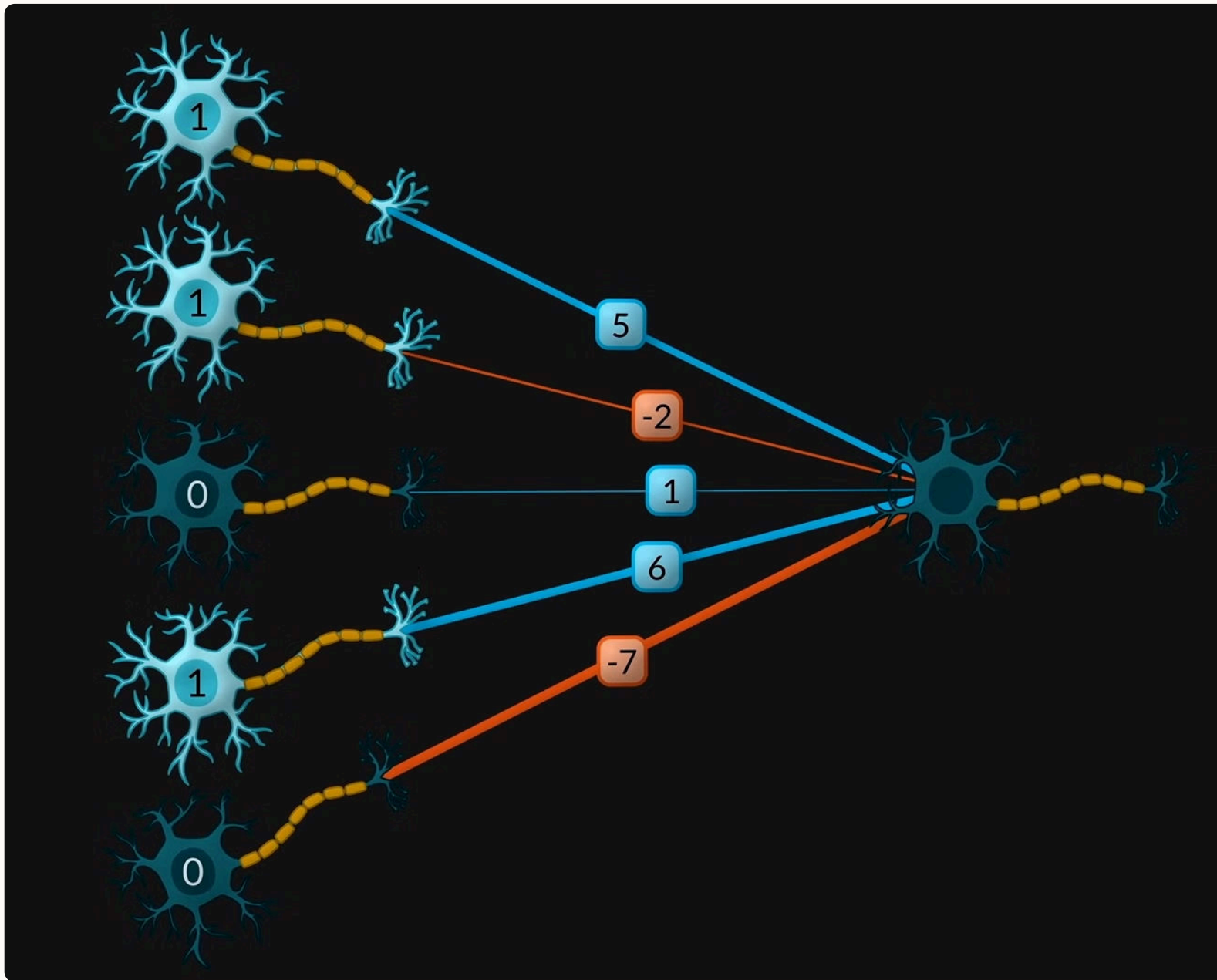
Умножение

$$I = U \cdot G$$

$$G = \frac{1}{R} \text{ — проводимость}$$

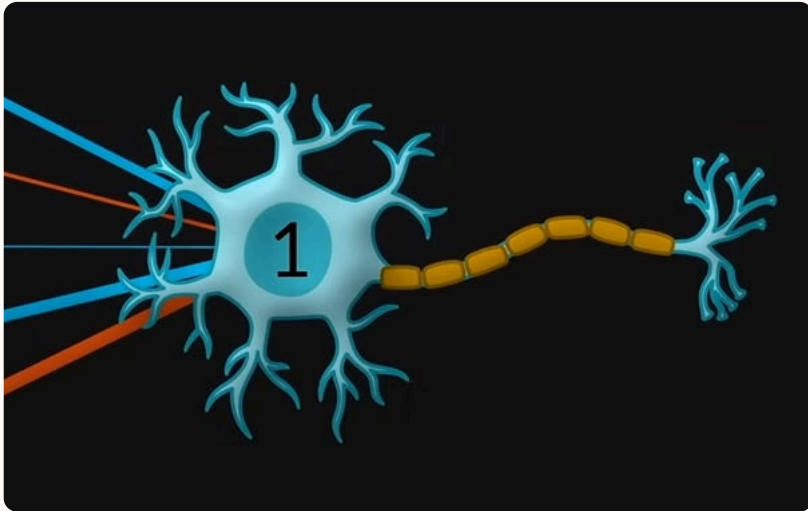


Нейрон простыми словами

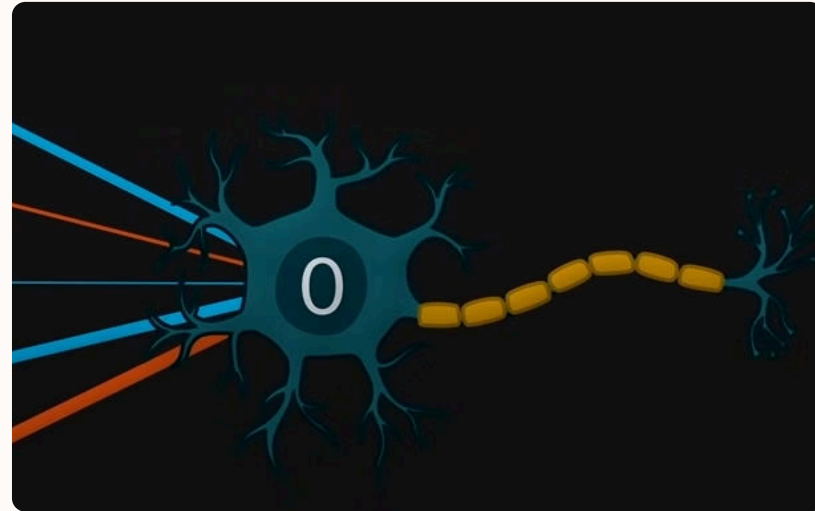


$$c = 0 \cdot 5 + 1 \cdot (-2) + 0 \cdot 1 + 1 \cdot 6 + 0 \cdot (-7)$$

Нейрон простыми словами

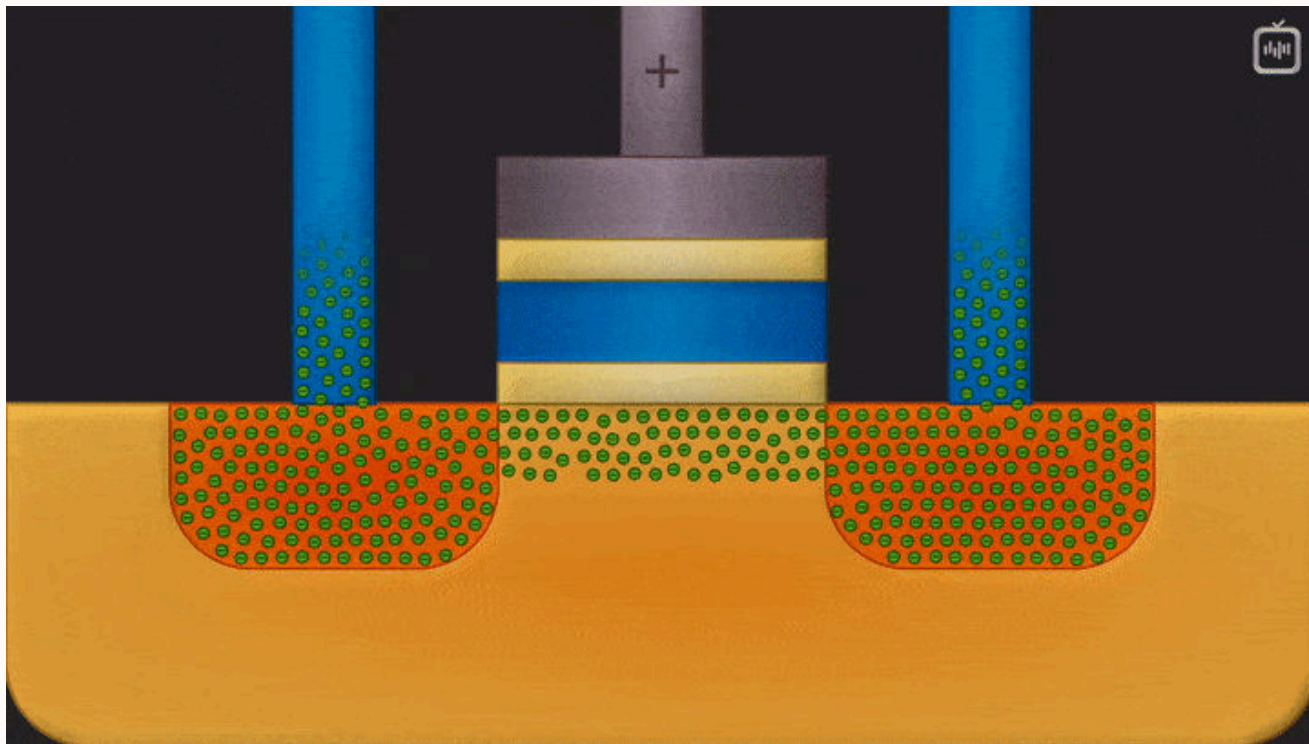


$\text{bias} < c$

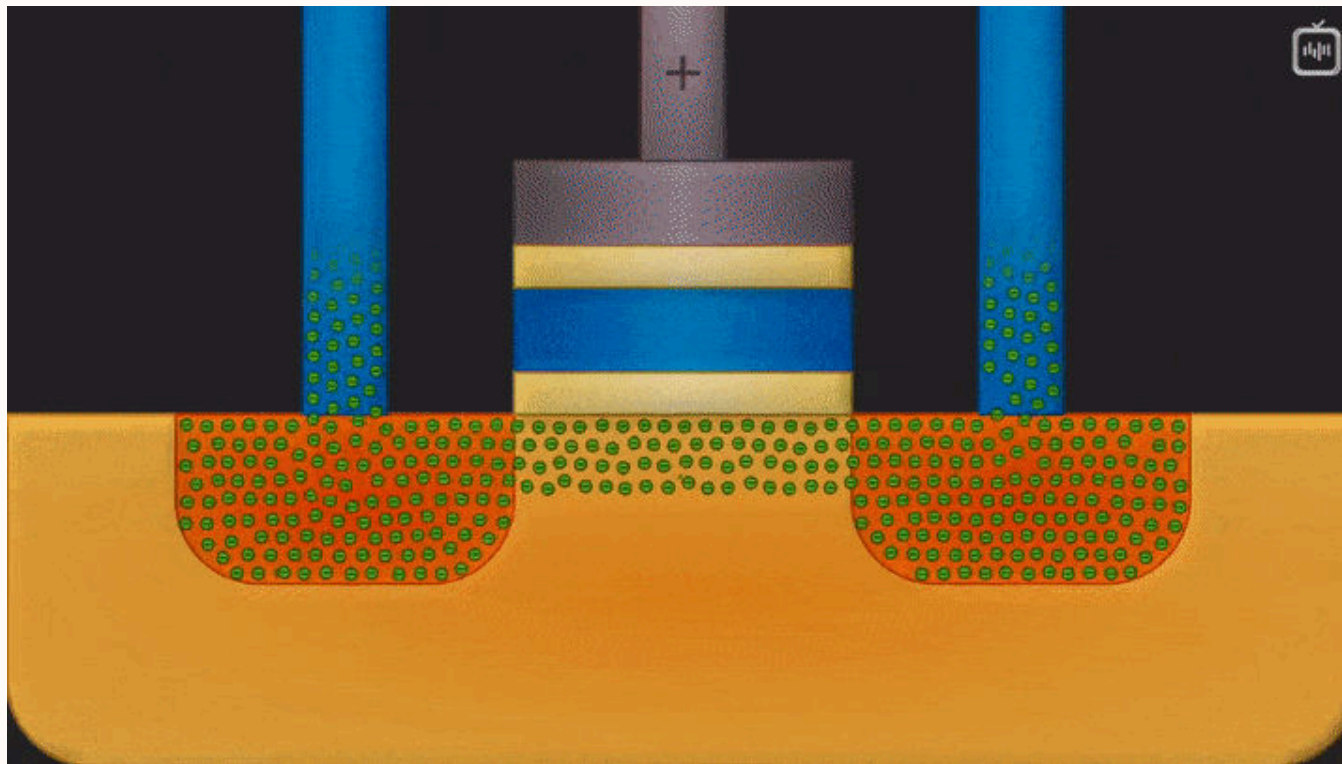


$\text{bias} > c$

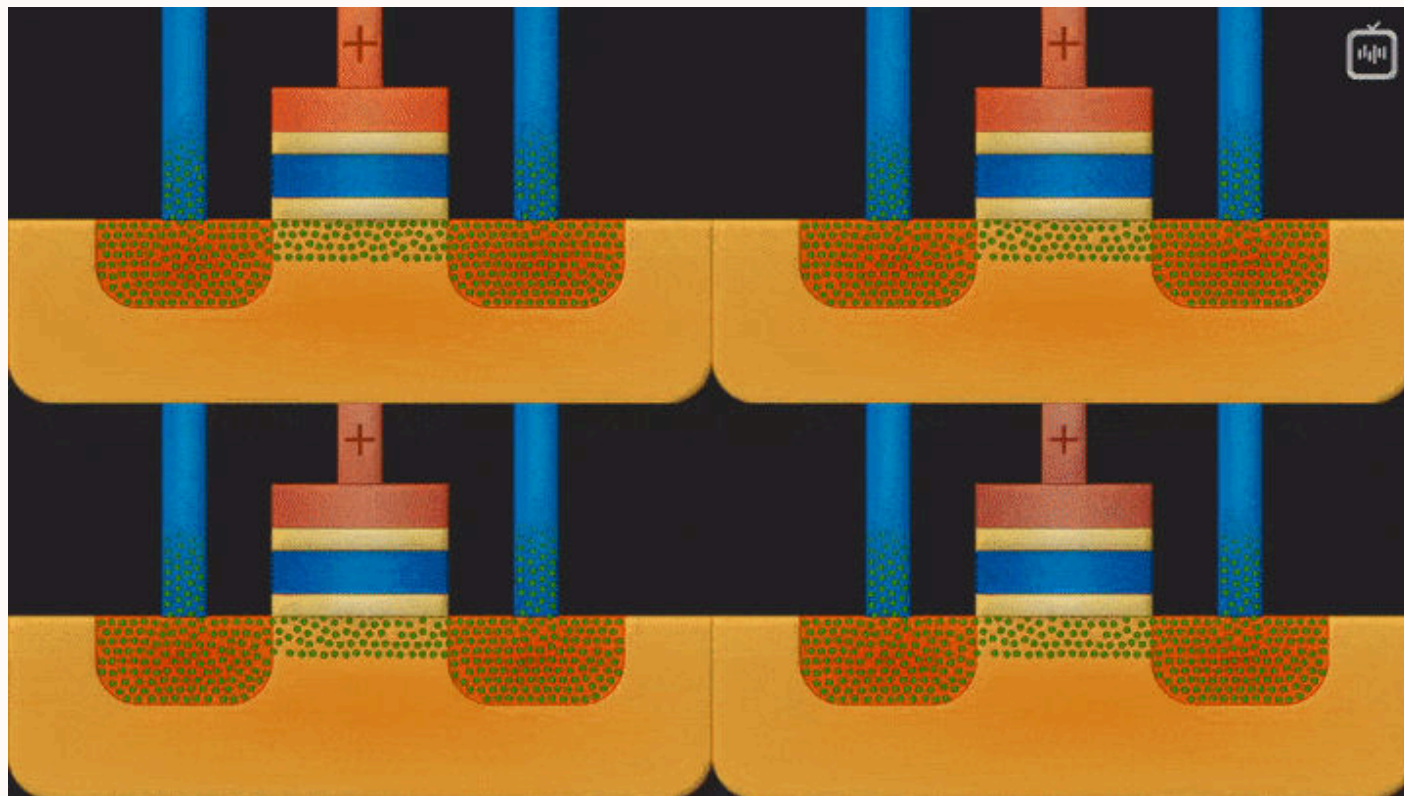
Иновация от Mythic



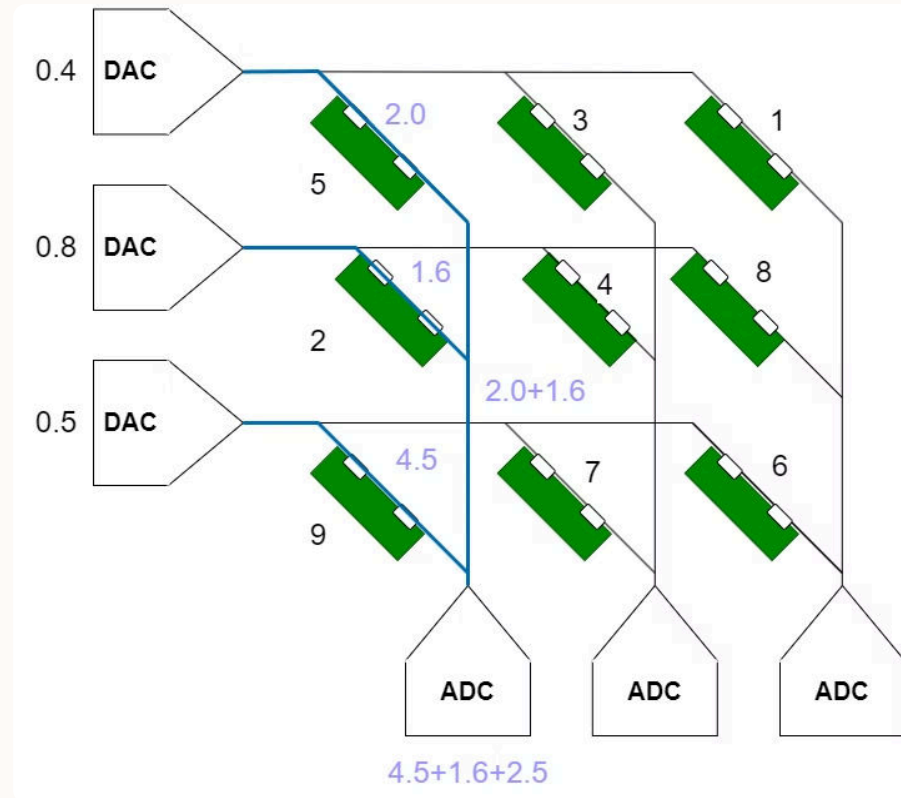
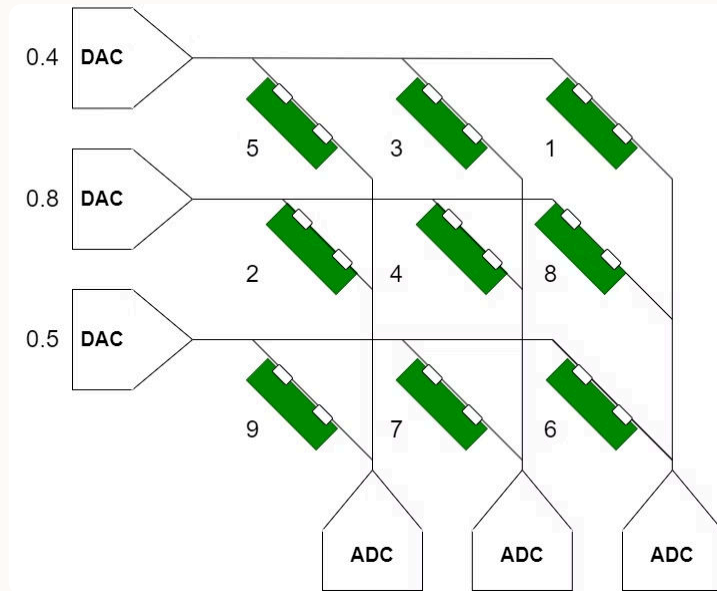
Иновация от Mythic



Иновация от Mythic



Вернемся к матрицам

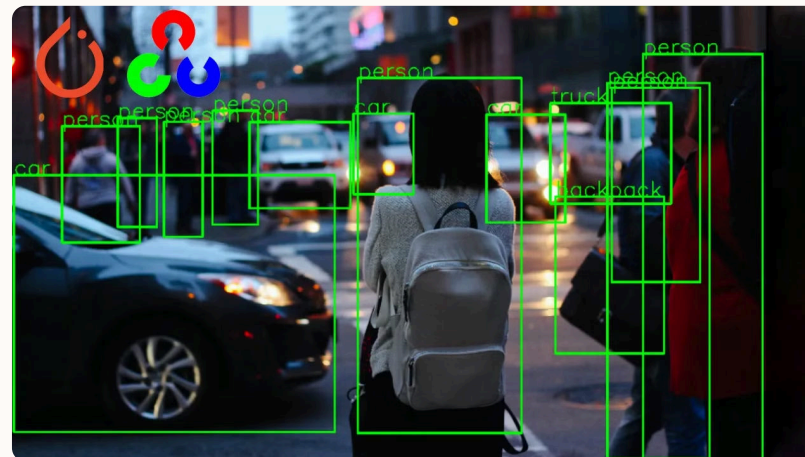


Какой результат?

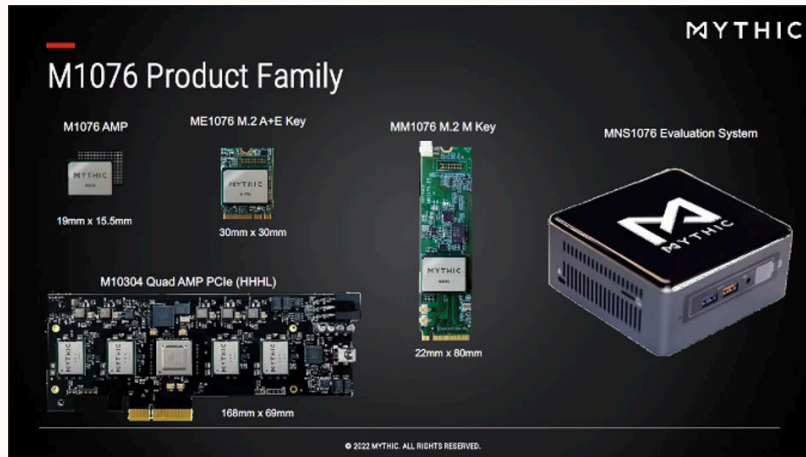
- Нет регистров, дополнительных вычислительных блоков - меньший размер чипа
- За один электрический импульс вычисления проходят во всех точках
- Увеличение количества матричных операций в единицу времени
- Потребление не больше обычных твердотельных накопителей

Потенциальные области применения

- Камеры видеонаблюдения с распознаванием лиц/жестов
- Производственные аппараты с анализом качества продукции
- Беспилотные системы



Сравнительный анализ (Mythic AI)



Mythic M10304

- 100 TOPS (INT8)
- 25 W
- ~150\$



NVIDIA Tesla A100

- 624-1248 TOPS (тензорное ядро для INT8)
- 300 W
- ~ 12000-30000\$

Сравнительный анализ (Mythic AI)

Плюсы

- Сравнительно маленькое потребление энергии
- Более выгодная цена за 1 TOPS
- Небольшой размер
- Доступная стоимость
- Эффективные вычисления матричных операций

Минусы

- Очень узкий спектр решаемых задач
- Не подходит для обучения нейронных сетей

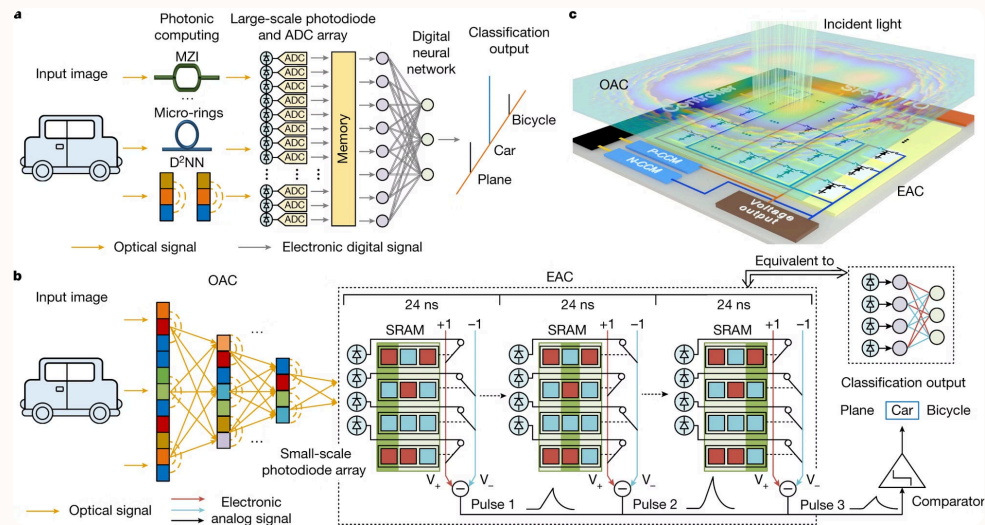
Итоги и замечания

Аналоговые чипы от Mythic AI показали хорошие результаты в задачах с искусственным интеллектом. Однако, в массовое производство чип так и не был выпущен, а его характеристики достоверны с точностью до слов главы компании Mythic.

Чип ACCEL

Фотоэлектронный чип, созданный командой ученых университета Цинхуа (All-analog Chip Combining Electronic and Light Computing)

Принцип работы



Сравнительный анализ (ACCEL)

Reference	Method	Energy efficiency (TOPS/W)	Computing speed (TOPS)	Processing latency (ns/frame)	Hardware implementation
Nature photonics (2021) ⁵ (claimed in paper)	Diffraction processing unit	1.58	240.1	2.07×10^7 (10-class 28 × 28 MNIST)	DMD + SLM + sCMOS camera (not chip)
Nature photonics (2021) ⁶	Diffraction processing unit (LeNet- equivalent)	0.000004	0.0006	2.07×10^7 (10-class 28 × 28 MNIST)	DMD + SLM + sCMOS camera (not chip)
Nature (2021) ² (claimed in paper)	Integrated microcomb accelerator	1.27	11.3	75.6 (10-class 28 × 28 MNIST)	Micro-ring resonator + MZM + SMF + PD
Nature (2021) ⁴ (claimed in paper)	Integrated photonic tensor core	0.4	4	8.1×10^3 (10-class 28 × 28 MNIST)	Si ₃ N ₄ micro-resonator + PCM (Ge ₂ Sb ₂ Te ₅) + PD
Nature (2022) ⁵ (claimed in paper)	Photo-electronic classification chip	0.07	0.27	0.57 (4-class 5 × 6 handwritten letters)	PIN attenuator + SiGe PD + micro-ring modulator
Google ⁴⁹ (claimed in paper)	TPU	2.30	92	NA	28 nm CMOS chip (digital)
NVIDIA ³³ (claimed on official website)	GPU A100 (parallel stream)	0.52	156	NA	7 nm CMOS chip (digital)
NVIDIA ³³ (experimentally measured)	GPU A100 (single stream LeNet)	0.001	0.08	2.6×10^5 (3-class 256 × 256 ImageNet)	7 nm CMOS chip (digital)
ISSCC (2022) ⁵⁰ (claimed in paper)	Computing-on-memory- boundary	32.9	0.97	1.3×10^6 (10-class 32 × 32 CIFAR-10)	28 nm CMOS chip (analog)
Proposed (experimentally measured)	ACCEL	7.48×10^4	4.55×10^3	72 (3-class 256 × 256 ImageNet)	SiO ₂ + 180 nm CMOS chip (analog)
Proposed (experimentally measured)	ACCEL (LeNet- equivalent)	4.95×10^3	301.39	72 (3-class 256 × 256 ImageNet)	SiO ₂ + 180 nm CMOS chip (analog)
Proposed (experimentally measured)	EAC only	1.40	0.09	72 (3-class 256 × 256 ImageNet)	180 nm CMOS chip (analog)

Сравнительный анализ (ACCEL)

Reference	Method	Energy efficiency (TOPS/W)	Computing speed (TOPS)
NVIDIA (claimed on official website)	GPU A100 (parallel stream)	0.52	156
NVIDIA (experimentally measured)	GPU A100 (single stream LeNet)	0.001	0.08
Proposed (experimentally measured)	ACCEL	$7.48 \cdot 10^4$	$4.55 \cdot 10^3$
Proposed (experimentally measured)	ACCEL (LeNet- equivalent)	$4.95 \cdot 10^3$	301.39

Сравнительный анализ (ACCEL)

Плюсы

- Чип имеет большой потенциал в массовом производстве
- Потрясающая производительность

Минусы

- Огромный тех-процесс на 180 нм
- Неопределенная цена
- Тесты проводились заинтересованными лицами
- Невозможность тренировки модели на данном чипе

Заключение

Аналоговые системы имеют большие перспективы в сфере искусственного интеллекта. Хотя все существующие решения не могут обучать модели самостоятельно, вполне возможно, что в скором времени станет выгодно обучать некоторую модель один раз и распространять ее на аналоговых системах.

Вопрос

Хранение и обработка взвешенной матрицы в твердотельном накопителе на аналоговой схеме
Mythic