

7. Архитектура сетей жизни vs Deep Learning

Итак из предыдущего раздела следуют, что ИТ уже десять лет как уперлись в стену - в невозможность эффективно наращивать производительность методом грубой силы, необходим переход к к массово параллельному и асинхронному программированию. Этот процесс на макроуровне давно идет и свидетельство этому - это развитая мировая экономика API, но основной рабочей лошадкой пока еще является автомат выполненный в архитектуре фон Неймана, хотя он уже и обвешан со всех сторон специализированными процессорами и графическими ускорителями.

Это значит, что в ИТ системах человечество все больше и больше будет переходить к архитектурам, которые присущи живым системам в которых ресурсы эффективно распределяются и концентрируются в узлах коммуникационной сети, обеспечивая необходимые производительность, надежность, качество и адаптивность системы, построенной в такой архитектуре. Примеры: API экономика, любой ее сервисный узел спроектированный на микросервисах, любая органическая или социальная система.

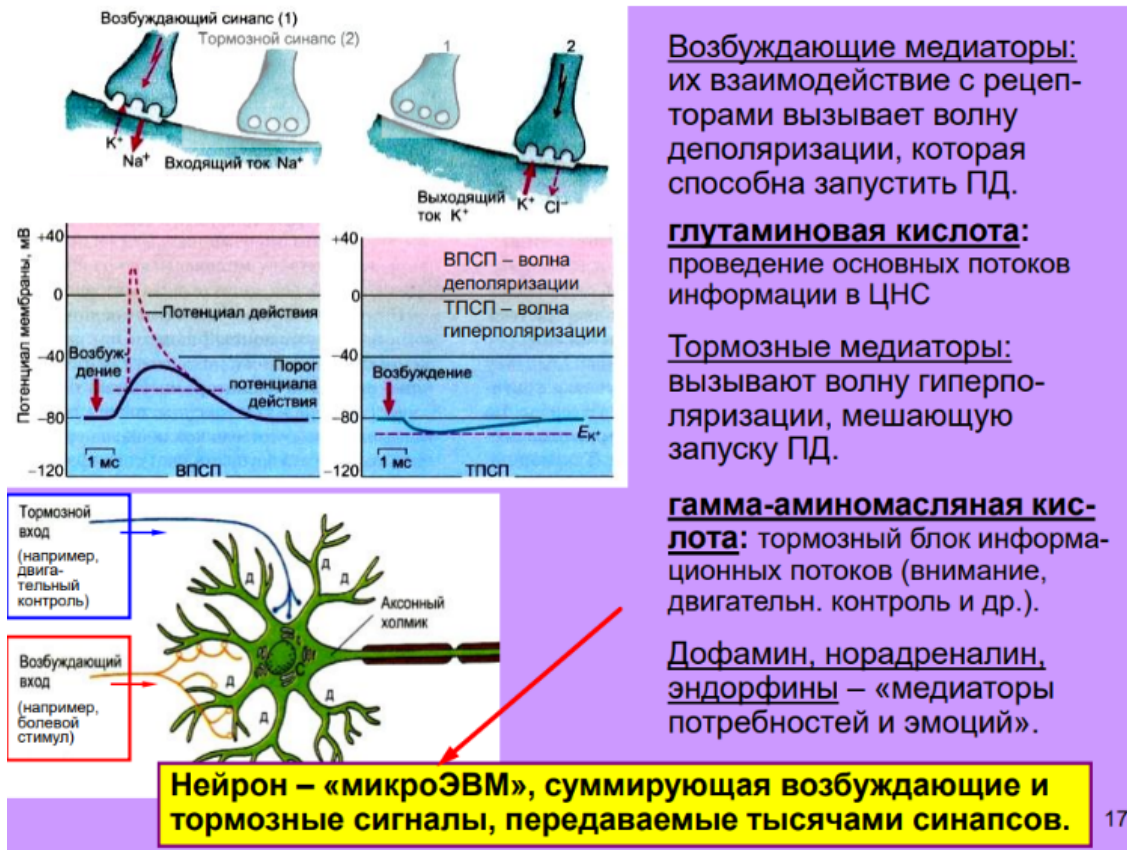
И здесь нам придется догонять Природу и еще многому у нее учиться. Чтобы показать это, сопоставим “наивысшее” достижение нашей цивилизации, так называемый искусственный интеллект, стоящий на технологиях Deep Learning и органические сети нашего мозга по различным критериям эффективности и по архитектуре.

Размер:

1. Размер кремниевого элемента: 10-20 атомов = 5- 10 нм (нм - нанометр 10⁻⁹)
2. Размер тела нейрона: 5-130 мкм = 5000 - 130 000 нм.

Казалось бы органические мозги не могут составить никакой конкуренции кремниевым. Но это не так, сопоставление размера кремниевого вентиля и размера нейрона некорректно, так как каждый нейрон является специализированным микропроцессором, а не отдельным его элементом. К тому же содержащим службу технического обслуживания и ремонта - не нужны никакие специалисты DevOps и инженеры технической поддержки. Сопоставлять с нейроном нужно специализированный процессор - например видеокарту или хотя бы одно его ядро.

На приведенном ниже рисунке проф. В. А. Дубынин очень верно схватил самую суть архитектуры мозга, за одним малым исключением - нейроном это не микроЭВМ (он не обладает памятью), нейрон - специализированный коммуникационный микропроцессор. Поэтому сопоставление в лоб размеров нейрона и кремниевого транзистора не корректно.



Поэтому сопоставим эффективный размер видеокарты приходящийся на одно ядро, например MSI GeForce RTX 3070 - ее размеры 14 см x 33.5 см x 3 слота

1. 10 см / 6144 ядер CUDA = 16 мкм
2. Размер тела нейрона: 5-130 мкм

Скорость:

1. Тактовая частота шины - 3.5 ггц, ширина канала - 64- 128 бит
2. Длина аксона - 0.5 мкм - 1.5 м, скорость нервного импульса 0.5-130 м/сек , длительность импульса 150-500 мкс.

Источник: [Аксон](#)

1. Ширина канала у кремния : за одну секунду по шине может быть передано 28-64 Гбайт данных.
2. Нейрон: 100-200 импульсов с числом потенциальных коммутаций на синапсах аксона до 20 тысяч.

Энергоэффективность:

1. Видеокарта: MSI GeForce RTX 3070 - 8 ГБ GDDR6X. 17.4 млрд транзисторов. ядра CUDA - 6 144. - потребляемая мощность - 170 Вт
2. Нейронов в мозгу 86 Млрд - потребляемая мощность - 10 Вт

1. 1 ядро CUDA - 27 милливатт (10⁻³)
2. 1 нейрон - 0.12 нановатт (10⁻⁹)

Источник: [Анатомия видеокарты - устройство и назначение](#)

Видим что кремниевые мозги недалеко ушли от органики в размерах и чудовищно проигрывают в энергоэффективности. С производительностью сложнее, так как принципы функционирования базовых элементов совершенно разные. Нейрон - это коммутатор сигналов, а ядро CUDA - программный автомат, который тасует данные в ячейках памяти. Напомню, что графические ускорители - это один из важнейших факторов в революции ИИ, произошедшей в 2000 годы - без них ИИ развивался бы намного медленнее и вряд ли нашел бы столь массовое применение, так как именно на видеоплатах производятся параллельные вычисления для обучения нейронной модели.

Версия #5

GRN создал 19 July 2022 09:18:13

GRN обновил 28 July 2022 18:47:22