The Future of Artificial Intelligence - Neuromorphic Computing
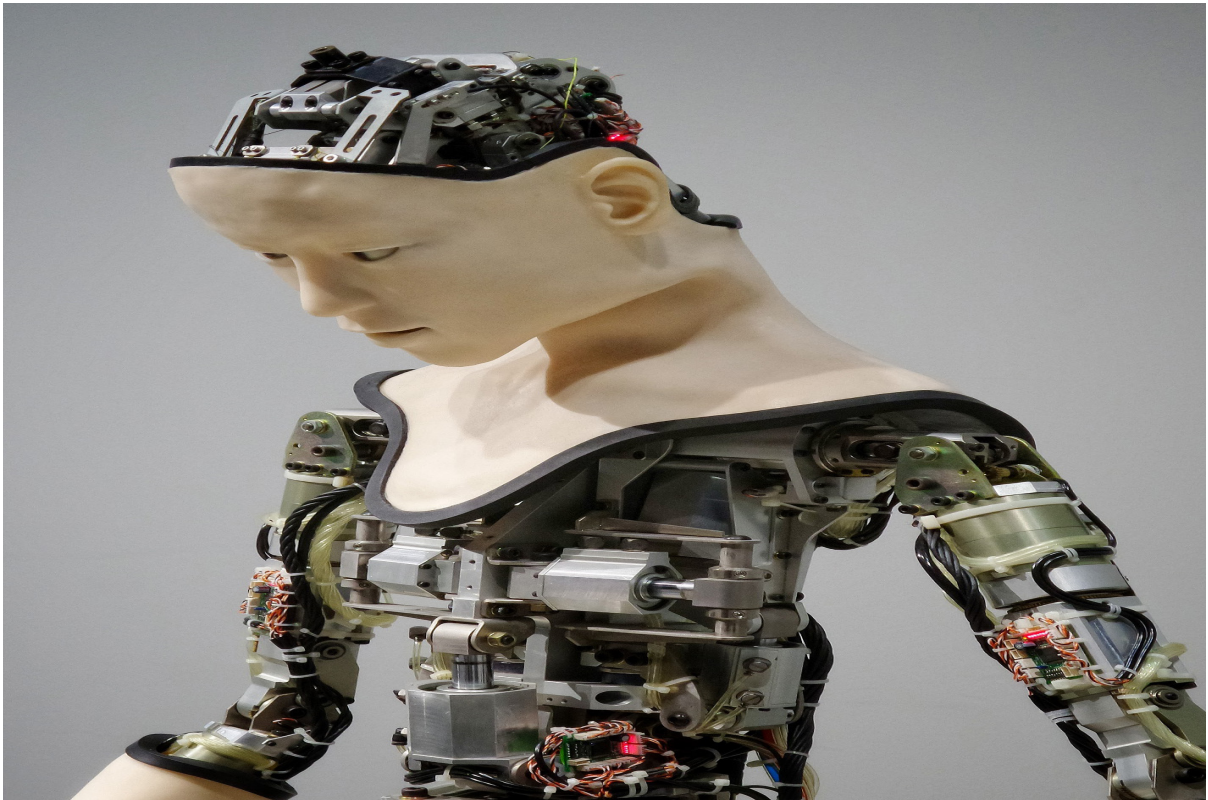

Photo by Franck V. on Unsplash

Everyone in the field of Artificial Intelligence knows what neural networks are. And most practitioners know the huge processing power and energy consumption needed to train pretty much any noteworthy neural network. That is to say, for the field to develop further, a new type of hardware is needed. Some experts consider that the quantum computer is that hardware. But even though it holds great promise, quantum computing is a technology that will take many decades to develop. Physics theories are not yet mature enough to enable the development of useful and cost-efficient devices. Neuromorphic computing, on the other hand, will take less time and resources to develop and will be very useful and cost-efficient, both regarding the cost of device development and the energy costs for processing power.

As most readers don't know what neuromorphic computing is, I will describe this new technology in this article, along with what it will enable us to do, in the "much-nearer-that-quantum" future. It all starts with the core electronic component that is used in neuromorphic chips, the memristor. **Memristor — The missing circuit element** Memristors have been proven to work in a similar way the brain's synapses work as they have a so-called plasticity. This plasticity can be used to create artificial structures inspired by the brain, to both process and memorize data. As it is well known, there are 3 fundamental passive electrical components: • **Capacitor** – that has the function to store electrical energy (in an electric field) and discharge this energy in the circuit, when needed;

• **Resistor** – is a passive device that creates resistance to the flow of electrons;

• **Inductor** – also called coil, choke or reactor, is a passive two-terminal electronic device that stores energy in a magnetic field, while current flows through it.

For many years, those were the only 3 fundamental passive electrical components, until the memristor came into the picture.

So what are they?
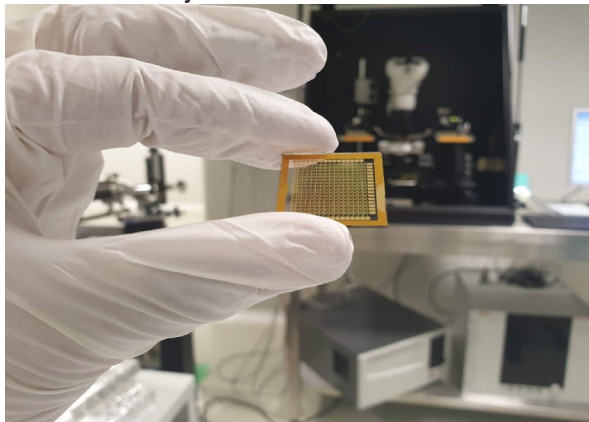
Memristor matrix created at CyberSwarm, Inc.

Memristors are a fourth class of electrical circuit, joining the components mentioned above, that exhibit their unique properties primarily at the nanoscale. Theoretically, memristors (a concatenation of memory and resistor) are a type of passive circuit element that maintains a relationship between the time integrals of current and voltage across a two-terminal element. Thus, the resistance of a memristor varies according to a devices' memristance function, allowing, via tiny read charges, to access a "history" (or memory) of the device.

All of this spelled in plain language means that this is a passive electrical component capable of "remembering" its past states (electrical resistance value), even when no energy passes through it. Memristors can also be used as both memory and processing units, an important feat that I will describe a bit later. The memristor was first discovered by Leon Chua, only in theory, in 1971. If you wish to read more

about it, you can find it in his original paper.

It wasn't until 2008 that a research team at HP labs first created a memristor, based on a thin film of Titanium Oxide. According to the team at HP Labs, memristors behave in the following way:

"The memristor's electrical resistance is not constant but depends on the history of current that had previously flowed through the device, i.e., its present resistance depends on how much electric charge has flowed in what direction through it in the past; the device remembers its history – the so-called non-volatility property. When the electric power supply is turned off, the memristor remembers its most recent resistance until it is turned on again."

Since then, various other materials have been experimented with by various companies, for the creation of memristors, which I won't enumerate.

But let's skip the boring stuff and get to the interesting Science-Fiction applications of the memristors!

Neuromorphic computing – or how to create a brain
Neuromorphic computing is nothing too new, as it was first coined in 1980 and it referred to analog circuits that mimic the neuro-biological architectures of the human brain.

If you are at least a bit into the hype of Deep Learning you know that it revolves around software-based algorithms and architectures that mimic, abstractly, the neural circuits of the brain. Deep Learning spreads more and more to all types of industries and the costs to train algorithms are huge.

One thing Deep Learning failed to mimic is the energy efficiency of the brain. Neuromorphic hardware can fix that.

Also, you can say goodbye to the famous Von Neumann bottleneck. In case you don't know what that is, it refers to the time it takes for the data coming from the memory of a device to reach the processing unit. This practically makes the processing unit wait (lose time) to get the data it needs to process.

Neuromorphic chips do not have a bottleneck, as all computations happen in memory. Memristors, which

are the base of neuromorphic chips, can be used as both memory units and computation units, similar to the way the brain does it. They are the first inorganic neurons.
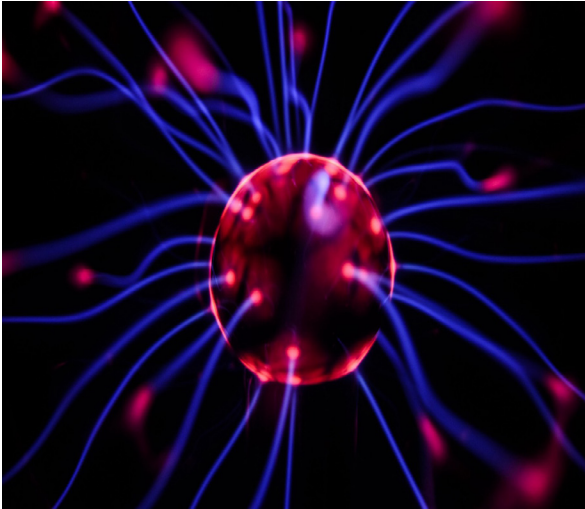

Photo by Josh Riemer on Unsplash

Why neuromorphic hardware is better for neural networks?

Neural networks primarily operate using real numbers (e.g. 2.0231, 0.242341, etc) to represent weights and other values inside of the neural network architecture. Those values, however, need to be transformed into binary, in the present computer architectures. That increases the number of operations needed to compute a neural network, both in training and deployment.

Neuromorphic hardware doesn't use binary, but it can use real values, under the form of electrical values like current and voltage. Here, the number 0.242341 is represented, for example, as 0.242341 volts. This happens directly inside the circuit, there is no binary value present. All of the calculations happen at the speed of the circuit.

Another factor that highly increases both the response speed and also the training speed of neural networks, based on neuromorphic hardware is the high parallelism of the calculations. One sure thing that is known about our brain is that it is highly par-allelized, millions of calculations happening at the same time, each second of our life. This is what neuromorphic chips can achieve.

All of those advantages come with a cherry on top: much lower energy consumption for training and deploying neural network algorithms.

Edge computing — powered by neuromorphic

As you may know, autonomous cars are based, mainly, on neural networks and 4/5 G technology. For a car to drive autonomously it has to be connected to a data center that analyses the data it receives from the car (usually it passes it to one or multiple convolutional neural networks) and then returns it to the car using 4/5 G technology. This produces latency, which can cost lives.

The advantage of the neuromorphic computing comes from the previously mentioned strong points of the technology; all processing can be done locally, inside the "brain" of the car (neuromorphic chip). This will reduce latency, reduce energy costs and increase the car's autonomy, while also improving a sensitive part of the car (as most data is processed locally) – cybersecurity.

It will also open many new capabilities for another growing industry: that of the Internet of Things. These devices are known to be very sensitive when it comes to energy consumption, as they have limited supplies of battery life. Small and very efficient neuromorphic chips could be used to give more utility to each device at low costs.

Conclusion

The Deep Learning industry needs a new type of hardware, neuromorphic hardware to be truly efficient. Neuromorphic chips show great promise to be the best fit for it and some companies are working towards developing this technology and the future of hardware artificial intelligence. Hopefully the industrial research in this area will increase and we will soon have hardware dedicated for neural networks. This would truly enable A.I. developers to change the world.