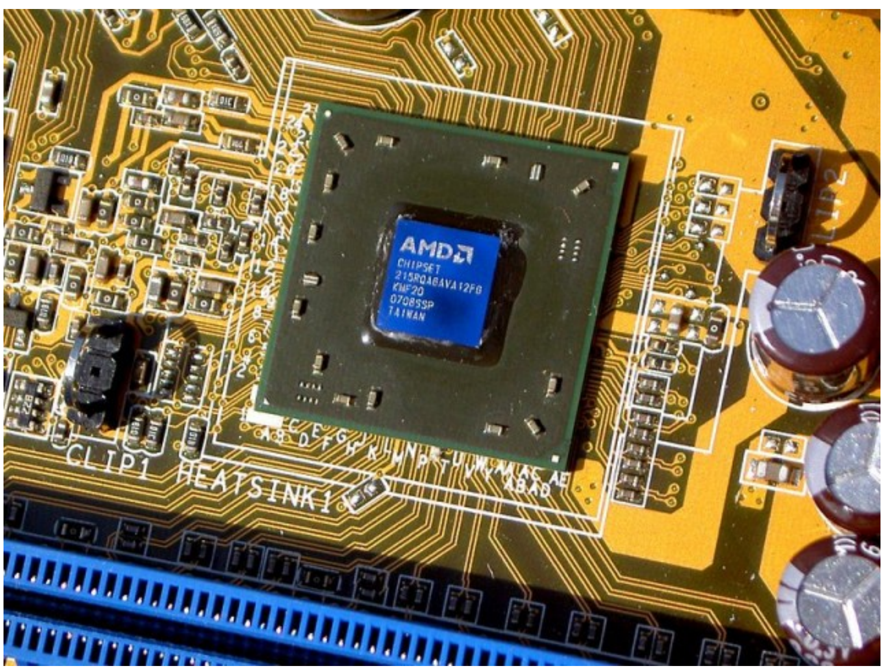
A BRAIN INSIDE A CHIP

Building machines which mimic the capacities of biological brains –and of the human brain in particular– is one of the main challenges of modern science. With the fast development of computer speed and power in the second half of the past century, this goal seemed, in the early days of artificial intelligence, simply a matter of time and resources. After all, computers were already able, in the 60s, to perform calculations at a speed many orders of magnitude above that of our best mathematicians, and they were also able to store and access vastly amounts of data (say, name and location of cities all over the world) at scales which are out of reach for normal humans.

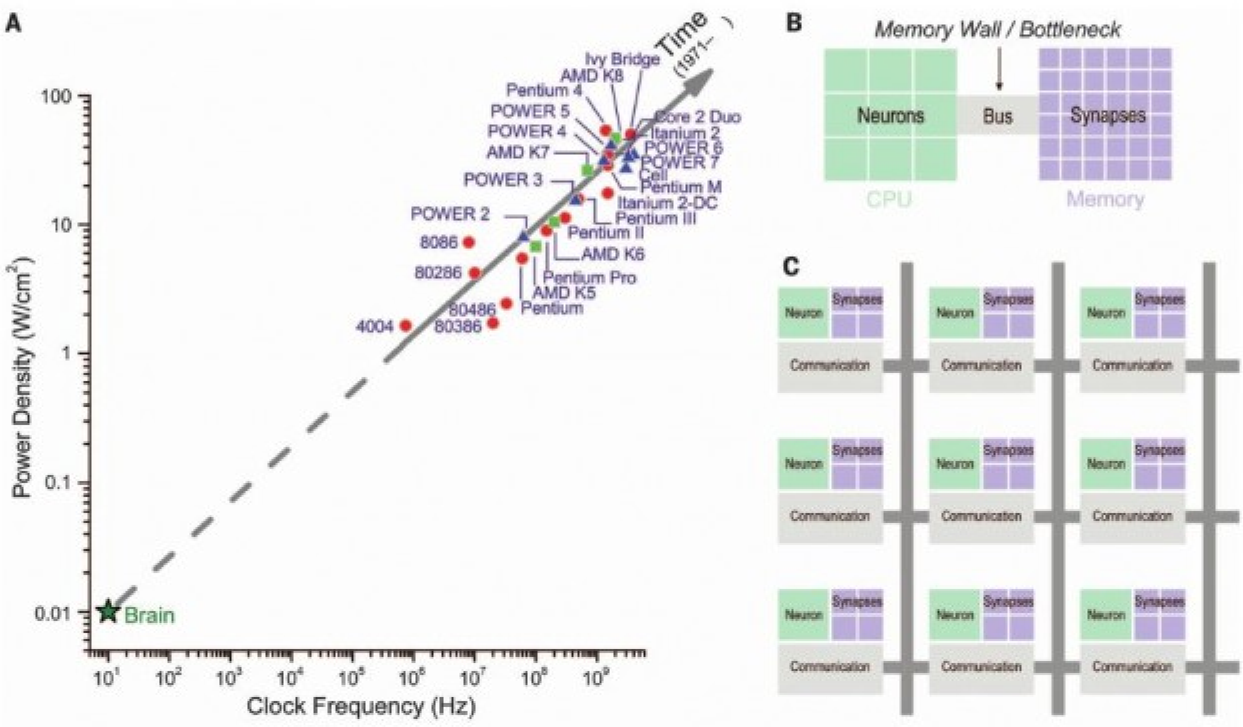


Microprocessors are the heart of our computers.

As soon as technology allowed it, researchers started to work on different approaches and implementations to allow a computer solve real-life tasks as well as we do. This list of problems include many vision-related tasks, such as object identification, face recognition, or contrast and rotation-invariant visual memory, although other relevant problems such as speech recognition, motor control for prosthetic arms, or value-based decision making are also worth mentioning. But ultimately, and in spite of its extreme efficiency in brute force numerical calculations, computers were –and more importantly, still are– far from being optimal to solve even a single one of the above mentioned 'real life' problems.

Researchers have suspected, for a few decades now, that the origin of such inefficiency is the poor performance of the von Neumann architecture when dealing with problems which require distributed and parallel computations. The von Neumann architecture –in which the current computer technology is based on – is fundamentally inefficient and nonscalable for representing massively interconnected neural networks. Such inefficiency is due to the separation between the external memory and the processor, in a way that imposes a bottleneck for computations and requires a lot of energy for data movement. It is only reasonable, therefore, that today's computers are no match for a real brain when dealing with object identification or speech recognition.

This scenario might change in the upcoming years, due to a new chip prototype invented by researchers at IBM leaded by Dr. Dharmendra Modha. The new chip architecture, which was presented recently in the journal *Science* [1], is inspired in the structure that real neurons adopt in the brain. Its goal is to capture some of the properties which make the brain an efficient task-solver, such as its distributed and parallel structure, or its event-driven nature. In the brain, neurons can be identified as information processing units (which, based on the input received over a period of time, decide whether to transmit the information with an action potential or not). On the other hand, synapses connecting these neurons are able to modify its effective strength in response to incoming information, which qualifies them as potential memory units in the brain. In the brain, therefore, information processing units and memory units are mixed together in the same structure – a neural network –, which prevents the formation of bottlenecks and facilitates parallelization. Moreover, real neural networks are event-driven, which means they only transmit information at very particular times (during the generation of action potentials) and this reduces the energy costs of computations considerably (see this for recent advances in brain energy consumption). In order to mimic these features, the novel neuro-inspired architecture developed by Modha and colleagues is based on tightly integrated memory, computation and communication modules that operate in parallel and communicate via an event-driven network



On the left, relation between power density and clock frequency of different computer processors; an estimation for the brain is shown for comparison. On the right, conceptual blueprint of von Neumann (top) and IBM's TrueNorth architecture.| Credit: Merolla et al (2014)

To test the properties and computational power of their design, the researchers built a prototype called *TrueNorth* based on the above architecture. It contains 5.4 billion transistors with 4096 cores, and integrates one million programmable spiking neurons and 256 million synapses. This prototype was used in an object classification task: the chip receives video information from a fixed camera, and it is trained to identify and track and categorize different moving objects (such as people walking down the street, cyclists, cars or trucks). During the task, the chip was found to consume 63 miliwatts on a 30-frame per second three-color video, which means it consumes about 150,000 times less energy than a general-purpose microprocessor, and almost 800 times less energy than a state-of-the-art multiprocessor neuromorphic chip. This massive decrease in energy consumption sets a new benchmark for future processors, and most likely will turns on the engines of a new era of neuro-inspired computers.

Source : http://mappingignorance.org/2014/10/06/brain-inside-chip/