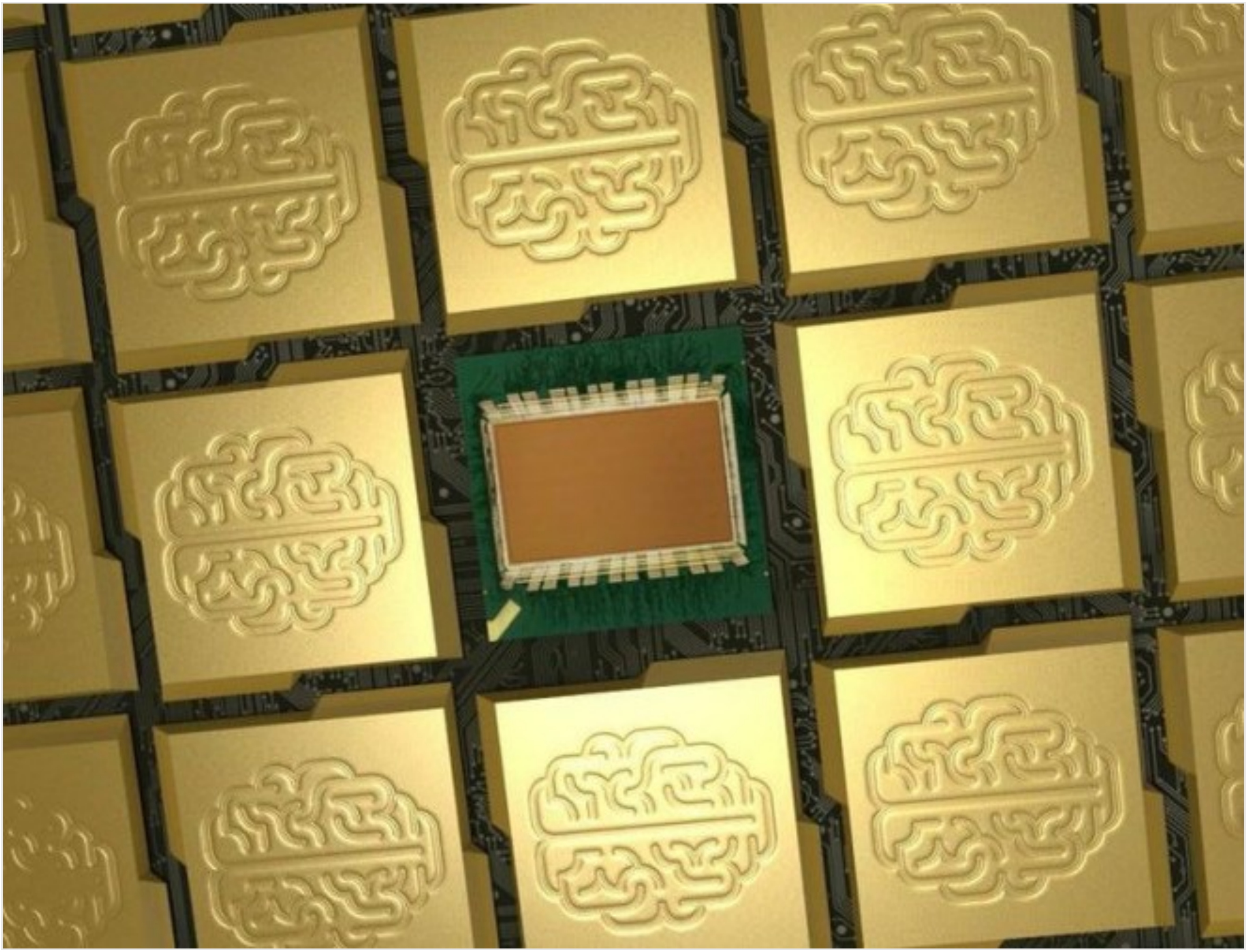
IBM Unveils a 'Brain-Like' Chip With 4,000 Processor Cores



IBM

The human brain is the world's most sophisticated computer, capable of learning new things on the fly, using very little data. It can recognize objects, understand speech, respond to change. Since the early days of digital technology, scientists have worked to build computers that were more like the three-pound organ inside your head.

Most efforts to mimic the brain have focused on software, but in recent years, some researchers have ramped up efforts to create neuro-inspired computer chips that process information in fundamentally different ways from traditional hardware. This includes an ambitious project inside tech giant IBM, and today, Big Blue released a research paper describing the latest fruits of these labors. With this paper, published in the academic journal *Science*, the company unveils what it calls TrueNorth, a custom-made "brain-like" chip that builds on a simpler experimental system the company released in 2011.

TrueNorth comes packed with 4,096 processor cores, and it mimics one million human neurons and 256 million synapses, two of the fundamental biological building blocks that make up the human brain. IBM calls these "spiking neurons." What that means, essentially, is that the chip can encode data as patterns of pulses, which is similar to one of the many ways neuroscientists think the brain stores information.

"This is a really neat experiment in architecture," says Carver Mead, a professor emeritus of engineering and applied science at the California Institute of Technology who is often considered the granddaddy of "neuromorphic" hardware. "It's a fine first step." Traditional processors—like the CPUs at the heart of our computers and the GPUs that drive graphics and other math-heavy tasks—aren't good at encoding data in this brain-like way, he explains, and that's why IBM's chip could be useful. "Representing information with the timing of nerve pulses…that's just not been a thing that digital computers have had a way of dealing with in the past," Mead says.

TrueNorth comes packed with 4,096 processor cores, and it mimics one million human neurons and 256 million synapses.

IBM has already tested the chip's ability to drive common artificial intelligence tasks, including recognizing images, and according to the company, its neurons and synapses can handle such tasks with usual speed, using much less power than traditional off-the-shelf chips. When researchers challenged the thing with DARPA's NeoVision2 Tower dataset—which includes images taken from video recorded atop Stanford University's Hoover Tower—TrueNorth was able to recognize things like people, cyclists, cars, buses, and trucks with about 80 percent accuracy. What's more, when the researchers then fed TrueNorth streaming video at 30 frames per second, it only burned 63 mW of power as it processed the data in real time.

"There's no CPU. There's no GPU, no hybrid computer that can come within even a couple of orders of magnitude of where we are," says Dharmendra Modha, the man who oversees the project. "The chip is designed for real-time power efficiency." Nobody else, he claims, "can deliver this in real time at the vast scales we're talking about." The trick, he explains, is that you can tile the chips together easily to create a massive neural network. IBM created a 16-chip board just a few weeks ago that can process video in real time.

Both these chips and this board are just research prototypes, but IBM is already hawking the technology as something that will revolutionize everything from cloud services, supercomputers, and smartphone technology. It's "a new machine for a new era," says Modha. "We really think this is a new landmark in the history of brain-inspired computing." But others question whether this technology is all that different from current systems and what it can actually do.

Beyond von Neumann

IBM's chip research is part of the SyNAPSE project, short for Systems of Neuromorphic Adaptive Plastic Scalable Electronics, a massive effort from DARPA, the Defense Department's research arm, to create a brain-like hardware. The ultimate aim of the project—which has invested about $53 million since 2008 in IBM's project alone—is to create hardware that breaks the von Neumann paradigm, the standard way of building computers.

In a von Neumann computer, the storage and handling of data is divvied up between the machine's main memory and its central processing unit. To do their work, computers carry out a set of instructions, or programs, sequentially by shuttling data from memory (where it's stored) to the CPU (where it's crunched). Because the memory and CPU are separated, data needs to be transferred constantly.

Ever since, scientists have been trying to understand how the brain encodes and processes information with the hope that they can translate that into smarter computers.

This creates a bottleneck and requires lots of energy. There are ways around this, like using multi-core chips that can run tasks in parallel or storing things in cache—a special kind of memory that sits closer to the processor—but this buys you only so much speed-up and not so much in power. It also means that computers are never really working in real-time, says Mead, because of the communication roadblock.

We don't completely understand how the brain works. But in his seminal work, *The Computer and the Brain*, as John von Neumann himself said that brain is something fundamentally different from the computing architecture that bears his name, and ever since, scientists have been trying to understand how the brain encodes and processes information with the hope that they can translate that into smarter computers.

Neuromorphic chips developed by IBM and a handful of others don't separate the data-storage and data-crunching parts of the computer. Instead, they pack the memory, computation and communication parts into little modules that process information locally but can communicate with each other easily and quickly. This, IBM researchers say, resembles the circuits found in the brain, where the separation of computation and storage isn't as cut and dry, and it's what buys the thing added energy efficiency—arguably the chip's best selling point to date.

But Can It Learn?

But some question how novel the chip really is. "The good point about the architecture is that memory and computation are close. But again, if this does not scale to state-of-art problems, it will not be different from current systems where memory and computation are physically separated," says Eugenio Culurciello, a professor at Purdue University, who works on neuromorphic systems for vision and helped develop the NeuFlow platform in neural-net pioneer Yann LeCun's lab at NYU.

Big Blue envisions a world where its TrueNorth chip helps us find our way. But that may be years away. IBM

So far, it's unclear how well TrueNorth performs when it's put to the test on large-scale state-of-the-art problems like recognizing very many different types of objects. It seems to have performed well on a simple image detection and recognition tasks using used DARPA's NeoVision2 Tower dataset. But as some critics point out, that's only five categories of objects. The object recognition software used at Baidu and Google, for example, is trained on the ImageNet database, which boasts thousands of object categories. Modha says they started with NeoVision because it was a DARPA-mandated metric, but they are working on other datasets including ImageNet.

'If this does not scale to state-of-art problems, it will not be different from current systems where memory and computation are physically separated.'

Others say that in order to break with current computing paradigms, neurochips should learn. "It's definitely an achievement to make a chip of that scale…but I think the claims are a bit stretched because there is no learning happening on chip," says Nayaran Srinivasa, a researcher at HRL Laboratories who's working on similar technologies (also funded by SyNAPSE). "It's not brain-like in a lot of ways." While the implementation does happen on TrueNorth, all the learning happens off-line, on traditional computers. "The von Neumann component is doing all the 'brain' work, so in that sense it's not breaking any paradigm."

To be fair, most learning systems today rely heavily on off-line learning, whether they run on CPUs or faster, more power-hungry GPUs. That's because learning often requires reworking the algorithms and that's much harder to do on hardware because it's not as flexible. Still, IBM says on-chip learning is not something they're ruling out.

Critics say the technology still has very many tests to pass before it can supercharge data centers or power new breeds of intelligent phones, cameras, robots or Google Glass-like contraptions. To think that we're going to have brain-like computer chips in our hands soon would be "misleading," says LeCun, whose lab has worked on neural-net hardware for years. "I'm all in favor of building special-purpose chips for running neural nets. But I think people should build chips to implement algorithms that we know work at state of the art level," he says. "This avenue of research is not going to pan out for quite a while, if ever. They may get neural net accelerator chips in their smartphones soonish, but these chips won't look at all like the IBM chip. They will look more like modified GPUs."