

Neuromorphic Chips Are Destined for Deep Learning—or Obscurity

Researchers in this specialized field have hitched their wagon to deep learning's star

By **LEE GOMES** Posted 29 May 2017 | 19:00 GMT

People in the tech world talk of a technology “crossing the chasm” by making the leap from early adopters to the mass market. A case study in chasm crossing is now unfolding in neuromorphic computing.

The approach mimics the way neurons are connected and communicate in the human brain, and enthusiasts say neuromorphic chips can run on much less power than traditional CPUs. The problem, though, is proving that [neuromorphics](http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/analog-and-neuromorphic-chips-will-rule-robotic-age) (<http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/analog-and-neuromorphic-chips-will-rule-robotic-age>) can move from research labs to commercial applications. The field's leading researchers spoke frankly about that challenge at the [Neuro Inspired Computational Elements Workshop](https://www.src.org/calendar/e006125/) (<https://www.src.org/calendar/e006125/>), held in March at the IBM research facility at Almaden, Calif.

“There currently is a lot of hype about neuromorphic computing,” said [Steve Furber](http://www.computerhistory.org/fellowawards/hall/steve-furber/) (<http://www.computerhistory.org/fellowawards/hall/steve-furber/>), the researcher at the University of Manchester, in England, who heads the [SpiNNaker project](http://spectrum.ieee.org/computing/hardware/lowpower-chips-to-model-a-billion-neurons) (<http://spectrum.ieee.org/computing/hardware/lowpower-chips-to-model-a-billion-neurons>), a major neuromorphics effort. “It's true that neuromorphic systems exist, and you can get one and use one. But all of them have fairly small user bases, in universities or industrial research groups. All require fairly specialized knowledge. And there is currently no compelling demonstration of a high-volume application where neuromorphic outperforms the alternative.”

Other attendees gave their own candid analyses. Another prominent researcher, [Chris Eliasmith](#)



CAN WE
COPY THE
BRAIN?

Section 3:
Engineering
Cognition

[\(/static/special-report-can-we-copy-the-brain\)](#)

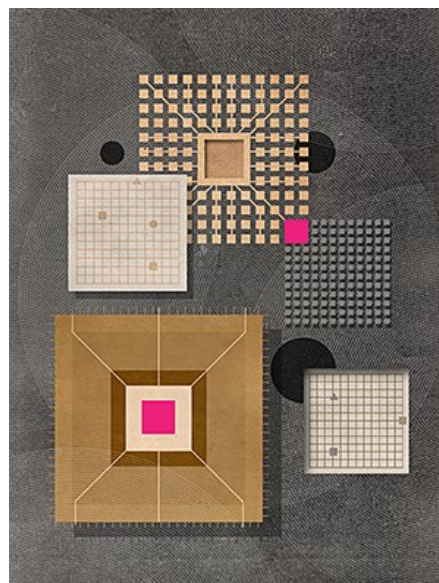


Illustration: Chad Hagen

(<http://arts.uwaterloo.ca/~celiasmi/>) of the University of Waterloo, in Ontario, Canada, said the field needs to meet the hype issue “head-on.” Given that neuromorphics has generated a great deal of excitement, Eliasmith doesn’t want to “fritter it away on toy problems”: A typical neuro morphic demonstration these days will show a system running a relatively simple artificial intelligence application. Rudimentary robots with neuromorphic chips have navigated down (<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&>

ved=oahUKEwjRgMrtv4jTAhWD6IMKHRW3BmoQF

%2Fwww.socsci.uci.edu%2F~jkrichma%2FHwu-arXiv-1611.01235v1.pdf&

usg=AFOjCNGhfo_no1NdWRzcWU9CRR3FONaEJg&cad=rja) a Colorado mountain trail and rolled over

squares (<https://www.youtube.com/watch?v=8c1Noq2K9cc>) of a specific color placed in a pattern on the

floor. The real test is for traditional companies to accept the “long more to do” before that happens. everyday engineering challenges, Eliasmith said, but there’s “long more to do” before that happens.

The basic building block of neuromorphic computing is what researchers call a **spiking neuron**, which plays a role analogous to what a logic gate does in traditional computing. In the central processing unit of your desktop, transistors are assembled into different types of logic gates—AND, OR, XOR, and the like—each of which evaluates two binary inputs. Then, based on those values and the gate’s type, each gate outputs either a 1 or a 0 to the next logic gate in line. All of them work in precise synchronization to the drumbeat of the chip’s master clock, mirroring the Boolean logic of the software it’s running.



**Baidu's Boffin Describes
Beijing's Homegrown Self-
Driving Car**

(cars-that-think/transportation/self-driving/baidus-boffin-describes-beijings-homegrown-self-driving-car-)

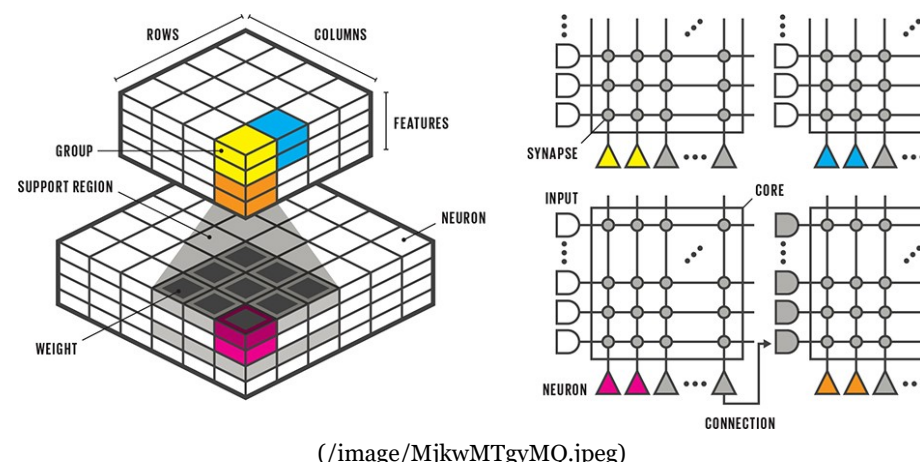


Illustration: James Provost

Tiny Spikes: Two layers within a neural network contain groups of “neurons” with similar functions, indicated by color [blue, yellow, orange, and pink] in the illustration on the left. In the graphic on the right, those neurons are mapped to spiking neurons in an IBM TrueNorth chip. The spiking neurons are connected by gridlike “synapses” to other neurons in the same core, and to a row of inputs. Those inputs can generate spikes, which are then processed by the neural network.

The spiking neuron is a different beast. Imagine a node sitting on a circuit and measuring whatever spikes—in the form of electrical pulses—are transmitted along the circuit. If a certain number of spikes occur within a certain period of time, the node is programmed to send along one or more new spikes of its own, the exact number depending on the design of the particular chip. Unlike the binary, 0-or-1 option of traditional CPUs, the responses to spikes can be weighted to a range of values, giving neuromorphics something of an analog flavor. The chips save on energy in large part because their neurons aren’t constantly firing, as occurs with traditional silicon technology, but instead become activated only when they receive a spiking signal.

“When we look at how neurons compute in the brain, there are concrete things we can learn”

A neuromorphic system connects these spiking neurons into complex networks, often according to a task-specific layout that programmers have worked out in advance. In a network designed for image recognition, for example, certain connections between neurons take on certain weights, and the way spikes travel between these neurons with their respective weights can be made to represent different objects. If one pattern of spikes appears at the output, programmers would know the image is of a cat; another pattern of

spikes would indicate the image is of a chair.

Within neuromorphics, each research group has come up with its own design to make this possible. IBM’s DARPA-funded [TrueNorth neuromorphic chip](http://science.sciencemag.org/content/345/6197/668) (<http://science.sciencemag.org/content/345/6197/668>), for example, does its spiking in custom hardware, while Furber’s [SpINNAker](http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6750072) (<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6750072>) (Spiking Neural Network

Architecture) relies on software running on the ARM processors that he helped develop.

In the early days, there was no consensus on what neuromorphic systems would actually do, except to somehow be useful in brain research. In truth, spiking chips were something of a solution looking for a problem. Help, though, arrived unexpectedly from an entirely different part of the computing world.

Starting in the 1990s, artificial intelligence researchers made a number of theoretical advances involving the design of the “neural networks” that had been used for decades for computational problem solving, though with limited success. Emre Neftci (<http://nmi-lab.org/people/>), with the University of California, Irvine’s Neuromorphic Machine Intelligence Lab, said that when combined with faster silicon chips, these new, improved neural networks allowed computers to make dramatic advances in classic computing problems, such as image recognition.

This new breed of computing tools used what’s come to be called deep learning, and in the past few years, deep learning has basically taken over the computer industry. Members of the neuromorphics research community soon discovered that they could take a deep-learning network and run it on their new style of hardware. And they could take advantage of the technology’s power efficiency: The TrueNorth chip (<http://spectrum.ieee.org/computing/hardware/how-ibm-got-brainlike-efficiency-from-the-truenorth-chip>), which is the size of a postage stamp and holds a million “neurons,” is designed to use a tiny fraction of the power of a standard processor.

Those power savings, say neuromorphics boosters, will take deep learning to places it couldn’t previously go, such as inside a mobile phone, and into the world’s hottest technology market. Today, deep learning enables many of the most widely used mobile features, such as the speech recognition required when you ask Siri a question. But the actual processing occurs on giant servers in the cloud, for lack of sufficient computing horsepower on the device. With neuromorphics on board, say its supporters, everything could be computed locally.

Which means that neuromorphic computing has, to a considerable degree, hitched its wagon to deep learning’s star. When IBM wanted to show off a killer app for its TrueNorth chip, it ran a deep neural network (<http://spectrum.ieee.org/tech-talk/computing/hardware/ibms-braininspired-chip-tested-on-deep-learning>) that classified images. Much of the neuromorphics community now defines success as being able to supply extremely power-efficient chips for deep learning, first for big server farms such as those run by Google, and later for mobile phones and other small, power-sensitive applications. The former is considered the easier engineering challenge, and neuromorphics optimists say commercial products for server farms could show up in as few as two years.

Unfortunately for neuromorphics, just about everyone else in the semiconductor industry—including big players like Intel (<https://software.intel.com/en-us/ai/deep-learning>) and Nvidia (<https://developer.nvidia.com/deep-learning>)—also wants in on the deep-learning market. And that market might turn out to be one of the rare cases in which the incumbents, rather than the innovators, have the strategic advantage. That’s because deep learning, arguably the most advanced software on the planet, generally runs on extremely simple hardware.

“The neuromorphic

Karl Freund

(<http://www.moorinsightsstrategy.com/karl-freund-biography/>), an analyst with Moor Insights & Strategy who specializes in deep learning, said the key bit of computation involved in running a

approaches are interesting scientifically, but they are nowhere close on accuracy”

In fact, most deep-learning systems use traditional silicon, especially the graphics coprocessors found in the video cards best known for powering video games. Graphics coprocessors can have thousands of cores, all working in tandem, and the more cores there are, the more efficient the deep-learning network.

So chip companies are bringing out deep-learning chips that are made out of very simple, traditional components, optimized to use as little power as possible. (That’s true of Google’s **Tensor Processing Unit** (<http://spectrum.ieee.org/tech-talk/computing/hardware/google-details-tensor-chip-powers>), the chip the search company announced last year in connection with its own deep-learning efforts.) Put differently, neuromorphics’ main competition as the platform of choice for deep learning is an advanced generation of what are essentially “vanilla” silicon chips.

Some companies on the vanilla side of this argument deny that neuromorphic systems have an edge in power efficiency. William J. Dally (http://cva.stanford.edu/billd_webpage_new.html), a Stanford electrical engineering professor and chief scientist at Nvidia, said that the demonstrations performed with TrueNorth used a very early version of deep learning, one with much less accuracy than is possible with more recent systems. When accuracy is taken into account, he said, any energy advantage of neuromorphics disappears.

“People who do conventional neural networks get results and win the competitions,” Dally said. “The neuromorphic approaches are interesting scientifically, but they are nowhere close on accuracy.”

Indeed, researchers have yet to figure out simple ways to get neuromorphic systems to run the huge variety of deep-learning networks that have been developed on conventional chips. **Brian Van Essen** (<https://people.llnl.gov/vanessen1>), at the Center for Applied Scientific Computing at the Lawrence Livermore National Laboratory, said his group has been able to get neural networks to run on TrueNorth but that the task of picking the right network and then successfully porting it over remains “a challenge.” Other researchers say the most advanced deep-learning systems require more neurons, with more possible interconnections, than current neuromorphic technology can offer.

The neuromorphics community must tackle these problems with a small pool of talent. The March conference, the field’s flagship event, attracted only a few hundred people; meetings associated with deep learning usually draw many thousands. IBM, which declined to comment for this article, said last fall that TrueNorth, which debuted in 2014, is now running experiments and applications for more than 130 users at more than 40 universities and research centers.

By contrast, there is hardly a Web company or university computer department on the planet that isn’t doing something with deep learning on conventional chips. As a result, those conventional architectures have a robust suite of development tools, along with legions of engineers trained in their use— typical advantages of an incumbent technology with a large installed base. Getting the deep-learning community to switch to a new and unfamiliar way of doing things will prove extremely difficult unless neuromorphics can offer an unmistakable performance and power advantage.

deep-learning system—known as matrix multiplication (<https://www.khanacademy.org/math/precalculus/precac-matrices/multiplying-matrices-by-matrices/v/matrix-multiplication-intro>)—can easily be handled with 16-bit and even 8-bit CPU components, as opposed to the 32- and 64-bit circuits of an advanced desktop processor.

Again, that's a problem the neuromorphics community openly acknowledges. If the presentations at the March conference frequently referred to the challenges that lie ahead for the field, most of them also offered suggestions on how to overcome them.

The University of Waterloo's Eliasmith, for example, said that neuromorphics must progress on a number of fronts. One of them is building more-robust hardware, with more neurons and interconnections, to handle more-advanced deep-learning systems. Also needed, he said, are theoretical insights about the inherent strengths and weaknesses of neuromorphic systems, to better know how to use them most productively. To be sure, he still believes the technology can live up to expectations. "We have been seeing regular improvements, so I'm encouraged," Eliasmith said.

Still, the neuromorphics community might find that its current symbiotic relationship with deep learning comes with its own hazards. For all the recent successes of deep learning, plenty of experts still question how much of an advance it will turn out to be.



(/image/MjkwNDIoMA.jpeg)

Photo: University of Manchester

Building Blocks: The SpiNNaker project is constructing a machine with 50,000 of these specialized chips in hopes of creating a network of 1 billion "neurons."

Center for Theoretical Neuroscience, said neuromorphic technology may, on its own, someday bring about AI results more sophisticated than anything deep learning ever could. "When we look at how neurons compute in the brain, there are concrete things we can learn," he said. "Let's try to build chips that do the same thing, and see what we can leverage out of them."

The SpiNNaker project's Furber echoed those sentiments when asked to predict when neuromorphics would be able to produce low-power components that could be used in mobile phones. His estimate was five years—but he said he was only 80 percent confident in that prediction. He added, however, that he was far more certain that neuromorphics would play an important role in studying the brain, just as early proponents thought it might.

Deep learning clearly delivers superior results in applications such as pattern recognition, in which one picture is matched to another picture, or for language translation. It remains to be seen how far the technique will take researchers toward the holy grail of "generalized intelligence" (<http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/why-alphago-is-not-ai>), or the ability of a computer to have, like HAL 9000 in the film *2001: A Space Odyssey*, the reasoning and language skills of a human. Deep-learning pioneer Yann LeCun (<http://yann.lecun.com/>) compares AI research to driving in the fog. He says there is a chance that even armed with deep learning, AI might any day now crash into another brick wall.

That prospect caused some at the conference to suggest that neuromorphics researchers should persevere even if the technology doesn't deliver a home run for deep learning. Bruno Olshausen (<http://redwood.berkeley.edu/bruno/>), director of the University of California, Berkeley's Redwood

However, there is a meta-issue hovering over the neuromorphics community: Researchers don't know whether the spiking behavior they are mimicking in the brain is central to the way the mind works, or merely one of its many accidental by-products. Indeed, the surest way to start an argument with a neuromorphics researcher is to suggest that we don't really know enough about how the brain works to have any business trying to copy it in silicon. The usual response you'll get is that while we certainly don't know everything, we clearly know enough to start.

It has often been noted that progress in aviation was made only after inventors stopped trying to copy the flapping wings of birds and instead discovered—and then harnessed—basic forces, such as thrust and lift. The knock against neuromorphic computing is that it's stuck at the level of mimicking flapping wings, an accusation the neuromorphics side obviously rejects. Depending on who is right, the field will either take flight and soar over the chasm, or drop into obscurity.

Advertisement

About the Author

Freelance journalist [Lee Gomes](https://www.linkedin.com/in/lee-gomes-5850b712) (<https://www.linkedin.com/in/lee-gomes-5850b712>) is a veteran technology reporter who has investigated claims surrounding wireless phone charging, self-driving cars, big data, and deep learning for *IEEE Spectrum*. In his spare time, he enjoys baking decadent French pastries and cakes.

This article appears in the June 2017 print issue as “The Neuromorphic Chip’s Make-or-Break Moment.”

SPECIAL REPORT: CAN WE COPY THE BRAIN?

[\(/static/special-report-can-we-copy-the-brain\)](/static/special-report-can-we-copy-the-brain)

PREVIOUS

**We Could Build an Artificial
Brain Right Now**

[\(/computing/hardware/we-could-build-](/computing/hardware/we-could-build-)

NEXT

**Watch This Robot Navigate
Like a Rat**

[\(/video/robotics/robotics-software](/video/robotics/robotics-software)



an-artificial-brain-right-now)

/watch-this-robot-navigate-like-a-rat)



Recommended For You



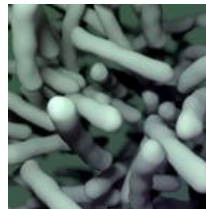
Startup Neurable Unveils the World's First Brain-Controlled VR Game

(/the-human-os/biomedical/bionics/brainy-startup-neurable-unveils-the-worlds-first-braincontrolled-vr-game)



Video Friday: More Boston Dynamics, Giant Fighting Robots, and ANYmal Quadruped

(/automaton/robotics/industrial-robots/video-friday-more-boston-dynamics-giant-fighting-robots-anymal-quadruped)



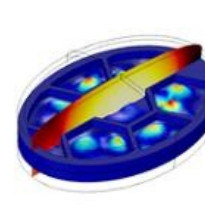
AI Makes Anthrax Bioterror Detection Easier

(/the-human-os/biomedical/imaging/ai-makes-anthrax-bioterror-detection-easier)



Teenage Whiz Kid Invents an AI System to Diagnose Her Grandfather's Eye Disease

(/the-human-os/biomedical/diagnostics/teenage-whiz-kid-invents-an-ai-system-to-diagnose-her-grandfathers-eye-disease)



Music to Your Ears: New Transducers Meet Electrostatic Headphones

(/computing/software/music-to-your-ears-new-transducers-meet-electrostatic-headphones)



Why Co-Parenting With Telepresence Robots Could Be a Fantastic Idea

(/automaton/robotics/home-robots/co-parenting-with-telepresence-robots)

