are generated by multiple thalamic inputs that have temporally different responses to the stimulus (Fig. 1). Thalamic inputs that respond slowly to visual stimuli generate slow responses in cortical regions, whereas those responding faster generate fast responses.

Lien and Scanziani's results, taken together with previous work[3–10], raise the interesting pos-sibility that cortical direction selectivity is gen-erated through a common mechanism — the convergence of temporally diverse thalamic inputs — in rodents, cats and primates. But as with all research, some questions remain open.

For instance, the authors focus their study on the middle layers of the visual cortex, which receive the bulk of the thalamic input[11]. As Lien and Scanziani show, many thalamic inputs in these middle cortical layers are not direction selective, but their combined activity is. It remains unclear whether thalamic inputs that target other cortical layers (or serve other functions) can encode direction selectivity through different mechanisms. For example, neurons in the superficial layers of the cor-tex might derive their direction selectivity from thalamic neurons that are themselves direction selective[12].

It is also known that thalamic inputs to the visual cortex are arranged by their receptive-field position — inputs that have receptive fields close to one another in the field of view are clus-tered together. However, it is not yet known whether the thalamic inputs are also arranged according to their temporal properties. If so, this could explain why spatial position and direction preference tend to change together in different neurons across the visual–cortical map[13].

Whatever the answers are, it is becoming increasingly clear that the visual cortex generates stimulus selectivity, such as preferences for direction and orientation, through thalamo–cortical convergence. Lien and Scanziani's work shows that this mechanism is better preserved across mammals than was previously thought. ■

Jose Manuel Alonso *is in the College of Optometry, State University of New York, New York, New York 10036, USA.*
e-mail: jalonso@sunyopt.edu

1. Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
2. Lien, A. D. & Scanziani, M. *Nature* **558**, 80–86 (2018).
3. Alonso, J.-M., Usrey, W. M. & Reid, R. C. *J. Neurosci.* **21**, 4002–4015 (2001).
4. Ferster, D., Chung, S. & Wheat, H. *Nature* **380**, 249–252 (1996).
5. Saul, A. B. & Humphrey, A. L. *J. Neurophysiol.* **64**, 206–224 (1990).
6. Saul, A. B. & Humphrey, A. L. *J. Neurophysiol.* **68**, 1190–1208 (1992).
7. Stanley, G. B. *et al. J. Neurosci.* **32**, 9073–9088 (2012).
8. Reid, R. C., Soodak, R. E. & Shapley, R. M. *J. Neurophysiol.* **66**, 505–529 (1991).
9. Livingstone, M. S. *Neuron* **20**, 509–526 (1998).
10. McLean, J. & Palmer, L. A. *Vision Res.* **29**, 675–679 (1989).
11. Lorente de No, R. In *Physiology of the Nervous System* (ed. Fulton, J.) 291–340 (Oxford Univ. Press, 1938).
12. Cruz-Martín, A. *et al. Nature* **507**, 358–361 (2014).
13. Kremkow, J., Jin, J., Wang, Y. & Alonso, J. M. *Nature* **533**, 52–57 (2016).

This article was published online on 23 May 2018.

ENGINEERING

Two artificial synapses are better than one

Emerging nanoelectronic devices could revolutionize artificial neural networks, but their hardware implementations lag behind those of their software counterparts. An approach has been developed that tips the scales in their favour. SEE ARTICLE P.60

GINA C. ADAM

Inspired by the brain's neural networks, scientists have for decades tried to construct electronic circuits that can process large amounts of data. However, it has been difficult to achieve energy-efficient implementations of artificial neurons and synapses (connections between neurons). On page 60, Ambrogio *et al.*[1] report an arti-ficial neural network containing more than 200,000 synapses that can classify complex collections of images. The authors' work dem-onstrates that hardware-based neural networks that use emerging nanoelectronic devices can perform as well as can software-based networks running on ordinary computers, while consuming much less power.

Artificial neural networks are not programmed in the same way as conventional computers. Just as humans learn from experi-ence, these networks acquire their functions from data obtained during a training process. Image classification, which involves learning and memory, requires thousands of artificial synapses. The states (electrical properties) of these synapses need to be programmed quickly and then retained for future network operation.

Nanoscale synaptic devices that have programmable electrical resistance, such

50 Years Ago

Reading aids for the blind have so far involved the use of intact sensory pathways and have progressed little beyond Braille and tape-recorded "talking-books". Both these systems are quite expensive … and both are slow in terms of information transfer to the reader … At a recent meeting of the Physiological Society, Brindley and Lewin demonstrated a device for stimulating the visual cortex of man directly … Essentially it consists of an array of radio receivers, encapsulated in silicone rubber and screwed to the skull … Activation of a receiver stimulated the cortex: transmission was in the form of a train of short (200 μs) pulses … it does at least seem feasible to transmit visual information directly to the central visual pathways of the recently blind.
From *Nature* 8 June 1968

100 Years Ago

It happened last week that about 1 lb. of fresh lamb was put into an oven at night in order that it might be cooked by morning on the "hay-box" principle. It was in a casserole, with a little water. Similar

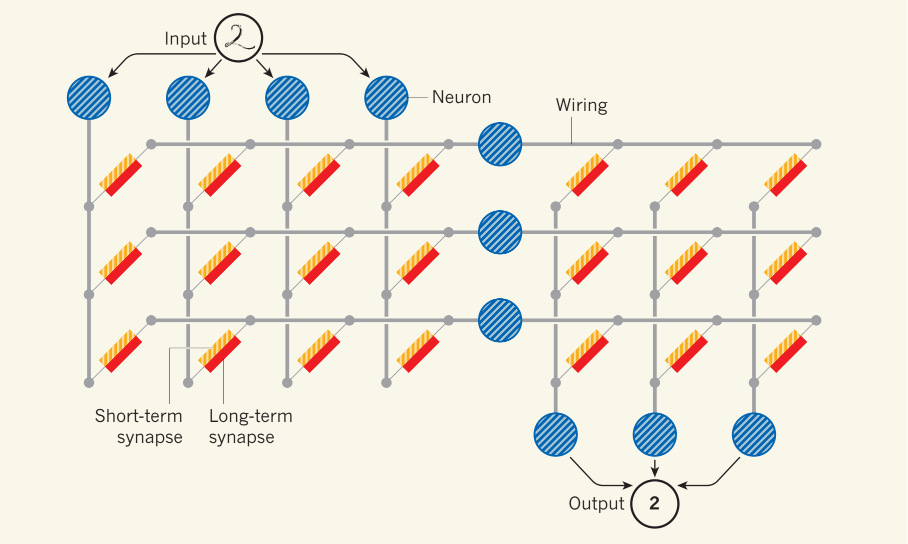

Figure 1 | An artificial neural network containing two types of synapse. Ambrogio *et al.*[1] report a hardware-based artificial neural network that is trained to classify complex images, such as handwritten numbers, with an accuracy similar to that of a software-based network. The network consists of artificial neurons linked by wires to two types of artificial synapse (connections between neurons). Short-term synapses (which can retain alterations in their synaptic state for milliseconds) are used regularly during network training, whereas long-term synapses (with state retention of years) are used mainly for memory. The long-term synapses are physical devices, whereas the neurons and short-term synapses are simulated computationally (indicated by hatching).

as phase-change-memory (PCM) devices, show promise because of their small physical size and excellent retention properties. PCM devices contain a material known as a chalcogenide glass, which can switch reversibly between an amorphous phase (of high resistance) and a crystalline phase (of low resistance). The device's resistance state is programmed by crystallizing part of the material using local heating produced by an applied voltage. This state is retained long after the voltage has been removed, and further programming can be achieved by crystallizing other parts of the material.

Unfortunately, PCM devices can be programmed in only one direction: from high to low resistance, by changing from low to high crystallinity. To achieve the desired resistance state with good precision, sequences of hundreds of voltage pulses are required. If the desired state is overshot, the chalcogenide glass must be completely reset to the amorphous phase and the step-by-step programming restarted. This shortcoming, combined with variations between devices caused by the manufacturing process, can slow or even prevent network training, as previous work by the group that performed the current study has shown[2]. As a result, the prototype networks that have been constructed using these devices[3,4] are impractical and have much lower image-classification accuracies than do software-based networks.

The breakthrough of Ambrogio and colleagues' work lies in a two-tier, bio-inspired approach. In biological neural networks, short-term changes in the states of synapses support a variety of computations, whereas long-term changes provide a platform for learning and memory[5]. For this reason, the authors' artificial neural network uses synaptic 'cells' that contain two types of synapse: short-term and long-term (Fig. 1).

The short-term synapses are used regularly during network training. They require only brief state retention, but fast and precise programming to the desired state. Such features are provided by an electronic switch called a transistor, which has a capacitor (a device for storing electric charge) attached to one of its electrodes, known as the gate[6]. The transistor's state is programmed by a fast voltage pulse applied to the gate. The capacitor maintains this voltage for a few milliseconds, providing brief state retention.

After the network has been trained on several thousand images and the short-term synapses have changed states substantially, the synaptic states are written into long-term synapses. The cycle is then repeated until all of the training images have been presented to the network. The long-term synapses are used for network operation after training is complete. They consist of PCM devices that have state-retention times of years, at the expense of tedious, energy-intensive programming.

An advantage of this technique is that the transfer of states from short- to long-term synapses can be done in electronic-circuit blocks separate from the network, while the network carries out other tasks. Moreover, although the authors' synaptic cells are more complicated in practice — containing one capacitor, two PCM devices and five transistors — they are still about half the size of artificial synapses used in other networks[6].

Ambrogio *et al.* tested their synaptic-cell approach using a fairly complex artificial neural network containing multiple layers of neurons and more than 200,000 PCM devices. The authors carried out classification tasks using three standard sets of images: greyscale handwritten numbers from the MNIST database[7], and colour images from the CIFAR-10 and CIFAR-100 databases[8]. The accuracies obtained were 98%, 88% and 68%, respectively. These results are strikingly similar to those obtained using TensorFlow, a leading neural-network software (see www.tensorflow.org).

Despite these impressive findings, a key limitation of the work is that only the PCM devices were actually fabricated; the other components of the synaptic cells and the neurons were simulated computationally. The authors took care to use accurate models that consider variations between transistors, and they proposed a method to minimize the impact of such variability on synaptic-cell performance. Most importantly, they carried out a detailed power assessment, and found that their proposed technology would consume about 100 times less power than current state-of-the-art networks, while providing a similar classification performance. Nevertheless, only a working hardware prototype will convince industry of the technology's performance and low-power advantages. Furthermore, the estimated power consumption is still a far cry from that of biological neural networks, leaving plenty of room for improvement.

However, Ambrogio and colleagues' work is more than a crucial stepping stone to the integration of PCM devices in neural-network hardware. It will also inspire device research, because it creates a need for nanoscale short-term synapses to replace the bulky transistor–capacitor ones. A wall in emerging memory technologies has been breached — networks based on these devices can work as well as do their software counterparts. This finding suggests that advances in artificial intelligence will not only continue, but also be accelerated by emerging hardware. ■

Gina C. Adam *is at the National Institute for Research and Development in Microtechnologies, Bucharest 077190, Romania.*
e-mail: gina.adam@imt.ro

1. Ambrogio, S. *et al. Nature* **558**, 60–67 (2018).
2. Burr, G. W. *et al. 2014 IEEE Int. Electron Devices Meet.* 29.5.1–29.5.4 (IEEE, 2014).
3. Gokmen, T. & Vlasov, Y. *Front. Neurosci.* **10**, 333 (2016).
4. Yu. S. *et al. 2015 IEEE Int. Electron Devices Meet.* 17.3.1–17.3.4 (IEEE, 2015).
5. Abbott, L. F. & Regehr, W. G. *Nature* **431**, 796–803 (2004).
6. Kim. S., Gokmen, T., Lee, H.-M. & Haensch, W. E. *2017 IEEE 60th Int. Midwest Symp. Circuits Systems* 422–425 (IEEE, 2017).
7. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. *Proc. IEEE* **86**, 2278–2324 (1998).
8. Krizhevsky, A. *Learning Multiple Layers of Features From Tiny Images* Ch. 3; www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf; https://www.cs.toronto.edu/~kriz/cifar.html (2009).