1. Rodriques, S. G. *et al. Science* **363**, 1463–1467 (2019).
2. Tang, F. *et al. Nature Methods* **6**, 377–382 (2009).
3. Macosko, E. Z. *et al. Cell* **161**, 1202–1214 (2015).
4. Klein, A. M. *et al. Cell* **161**, 1187–1201 (2015).
5. Zheng, G. X. *et al. Nature Commun.* **8,** 14049 (2017).
6. Shah, S., Lubeck, E., Zhou, W. & Cai, L. *Neuron* **92**, 342–357 (2016).
7. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. &
Zhuang, X. *Science* **348**, aaa6090 (2015).
8. Lee, J. H. *et al. Science* **343**, 1360–1363 (2014).
9. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G. C. *Nature Biotechnol.* **36**, 1183–1190 (2018).
10. Eng, C.-H. L. *et al. Nature* **568**, 235–239 (2019).

NANOTECHNOLOGY

# A role for optics in AI hardware

**Experiments show how an all-optical version of an artificial neural network — a type of artificial-intelligence system — could potentially deliver better energy efficiency can conventional computing approaches. SEE ARTICLE P.208**

**GEOFFREY W. BURR**

Optical fibres transmit data across the world in the form of light and are the backbone of modern telecommunications[1]. However, when such data need to be analysed, they get converted from light into electrons and are then processed using electronics. There was a time when optics was considered as the basis for a potential computing technology[2], but it became difficult for optical computing to compete with the rapid improvements made by its electronic counterpart. In the past few years, however, concern has been growing about the energy costs of computation. Therefore, optics is receiving attention again, both as a way to decrease energy requirements[3], and as a special-purpose hardware for accelerating artificial-intelligence algorithms such as deep neural networks (DNNs). On page 208, Feldmann *et al.*[4] report an intriguing advance towards all-optical implementations of such networks.

A DNN comprises many layers of artificial neurons and artificial synapses, which are connections between the neurons. The strengths of these connections are called weights and can be either positive, indicating neuronal excitation, or negative, implying inhibition. A DNN learns to perform tasks such as image recognition by varying its synaptic weights in a way that minimizes the difference between its actual output and the desired output.

Central processing units and other digital-based hardware accelerators[5] are typically used for DNN computations. A DNN can be trained using a known set of data, whereas an already trained DNN can be applied to unknown data in a task called inference. In either case, although the amount of computation is vast, the variety of operations is modest, because 'multiply–accumulate' operations dominate across the many synaptic weights and neuronal excitations.

DNNs are known to still work well when computational precision is low[5]. As a result, these networks represent an intriguing opportunity for unconventional computing techniques. For example, researchers are exploring DNN accelerators that are based on emerging non-volatile memory devices[6,7]. Such devices retain information even when their power source is switched off, and can offer improved speed and energy efficiency for DNNs through analog electronic computation.

Why not, therefore, also consider optics? Structures that direct light — whether they be an optical fibre for use in telecommunications or a waveguide patterned onto a photonic chip — can be packed with vast amounts of data. Inside such a waveguide, many wavelengths of light can propagate together, using a technique known as wavelength division multiplexing. Each wavelength can then be modulated (altered in such a way that it can carry information) at a rate that is limited by the available bandwidths associated with electronic-to-optical modulation and optical-to-electronic detection.

Structures called resonators enable individual wavelengths to be added to or removed from the waveguide, like wagons on a freight train. For example, micrometre-scale, ring-shaped (micro-ring) resonators can implement arrays of synaptic weights[8]. Such resonators can be modulated thermally[9], electro-optically[10,11] or, as in Feldmann and colleagues' work, through phase-change materials[12]. These materials can switch between an amorphous phase and a crystalline phase, which differ greatly in their ability to absorb light. Under ideal conditions, the resulting multiply–accumulate operations would require only a small amount of power.

Feldmann *et al.* present an all-optical neural network on a millimetre-scale photonic chip, in which there are no optical-to-electronic conversions within the network. Inputted data are electronically modulated onto different wavelengths for injection into the network, but after that has been performed, all the data stay on the chip. Both weight modulation and neuron integration are achieved using integrated phase-change materials; these are located on two types of micro-ring resonator, which have a synaptic or neuronal function.

Unmodulated light that is injected at the various operating wavelengths picks up the neuronal excitations that have accumulated in the phase-change material, and then passes them to the next layer of the network. Even without on-chip optical gain (a process in which a medium transfers energy to the light that is transmitted through it), this set-up could potentially be scaled up to larger networks. The authors demonstrate, on a small scale, both supervised and unsupervised learning — that is, training is achieved using labelled data, which is how DNNs learn, and using data without such labels, which is how humans tend to learn.

Because the weights are implemented by light absorption, negative weights require a large bias signal, which must not activate the phase-change material. An alternative approach[13] that can readily offer negative weights uses devices called Mach–Zehnder interferometers. In these devices, a single waveguide is split into two arms and then recombined; this causes the amount of transmitted light to depend on the difference in optical phase between the two paths. However, it might be challenging to combine this approach with wavelength division multiplexing, because the arms of each interferometer would need to introduce the appropriate phase difference for each wavelength.

Photonic DNNs still present substantial challenges. Their total power usage can be low in ideal situations, but thermo-optic power is frequently required to adjust and maintain the differences in optical phase in the arms of each Mach–Zehnder interferometer. Moreover, the total optical power that is injected into a system containing phase-change materials must be calibrated carefully, so that the materials respond to incoming signals exactly as intended. Although phase-change materials can also be used to adjust Mach–Zehnder phases, unavoidable cross-coupling between how strongly the materials absorb light and how much they slow it down poses a considerable complication.

Phase-change materials seem to be well suited for the non-volatile long-term storage of synaptic weights that are based on micro-ring resonators needing only infrequent adjustment. However, when used in the role of neuron, the speed of crystallization of such materials will limit the maximum rate at which neurons can be excited. Furthermore, the need to melt the materials to induce a full neuronal reset after every potential excitation event will rapidly consume the large, but finite,

switching endurance of the materials.

Conventional DNNs have grown large and now typically involve many thousands of neurons and millions of synapses. But photonic networks require waveguides that are spaced far from each other to prevent them from coupling, and that avoid sharp bends to prevent light from leaving the waveguide. Because crossing two waveguides introduces the risk of injecting undesired power into the wrong path, the 2D nature of a photonic chip presents a substantial design constraint.

Despite the long distances and large areas that are required for the implementation of photonic networks, fabrication of the key parts of each optical structure requires precision. This is because the waveguides and coupling regions — for instance, at the entrance and exit of each micro-ring resonator — must have the exact dimensions needed to obtain their desired performance. There are also limits to how small micro-ring resonators can

be made. Finally, the relatively weak optical effects offered by modulation techniques require long interaction regions to enable their limited impact on passing light to build to a noticeable level.

Advances such as those made in Feldmann and colleagues' study and by others[8,13] are encouraging for the future of the field. The development of readily available broadband on-chip gain would help considerably, as would techniques that can support independent and arbitrary operations on each piece of optically encoded data, without requiring vast areas of the photonic chip. Should scalable photonic neural accelerators offering high energy efficiencies eventually emerge, we might well look back on the work of Feldmann *et al.* and others in the field as important early glimpses of the technology's promise. ∎

**Geoffrey W. Burr** *is at IBM Research – Almaden, San Jose, California 95120, USA.*
*e-mail: gwburr@us.ibm.com*

1. Agrell, E. *et al. J. Opt.* **18**, 063002 (2016).
2. Karim, M. A. & Awwal, A. A. S. *Optical Computing: An Introduction* (Wiley, 1992).
3. Miller, D. A. B. *J. Lightwave Tech.* **35**, 346–396 (2017).
4. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. *Nature* **569**, 208–214 (2019).
5. Sze, V., Chen, Y.-H., Yang, T.-J. & Emer, J. S. *Proc. IEEE* **105**, 2295–2329 (2017).
6. Haensch, W., Gokmen, T. & Puri, R. *Proc. IEEE* **107**, 108–122 (2019).
7. Burr, G. W. *Nature Mach. Intell.* **1**, 10–11 (2019).
8. Tait, A. N. *et al. Sci. Rep.* **7**, 7430 (2017).
9. Padmaraju, K. & Bergman, K. *Nanophotonics* **3**, 269–281 (2014).
10. Soref, R. & Bennett, B. *IEEE J. Quantum Electron.* **23**, 123–129 (1987).
11. Xu, Q. & Lipson, M. *Optics Expr.* **15**, 924–929 (2007).
12. Ríos, C. *et al. Nature Photonics* **9**, 725–732 (2015).
13. Shen, Y. *et al. Nature Photonics* **11**, 441–446 (2017).

---

OBESITY

# Rural areas drive the global weight gain

**The global rise in the prevalence of obesity has been seen as an urban problem. A large-scale study challenges this view by showing that weight gain in rural areas is the main factor currently driving the obesity epidemic.** SEE LETTER P.260

**BARRY M. POPKIN**

Urbanization has been linked to increased overweight and obesity levels across populations[1]. However, evidence for this association has been based mainly on calculations of the body mass index (BMI) — the most frequently used tool for measuring overweight and obesity — at the time of study. The dynamics of BMI change in urban and rural areas have not been investigated separately. On page 260, the members of the NCD Risk Factor Collaboration[2] challenge the idea that general BMI trends are mainly a result of urbanization.

The global problem of overweight and obesity has been seen chiefly as an urban issue, partly because access to food services is much greater and easier in cities than in rural areas. City dwellers have an array of options for purchasing highly processed foods and beverages, which are high in salt, saturated fat and sugar, and which are often termed 'ultra-processed obesogenic foods'. Many low-income communities in urban areas consume predominantly ultra-processed foods and beverages sold at fast-food and small retail outlets, often because they live in so-called 'food deserts' — low-income areas where these

are the only available foods. Rural areas, on the other hand, have been seen as a different type of food desert, where people mainly consume produce from their own farms and gardens, and have less access to ultra-processed and packaged food.

Furthermore, the inhabitants of cities have better options for transport, greater access to smartphones and cable television, and more non-physical leisure opportunities than those living in rural areas. They are also more likely to have occupations that are not very physically demanding. All these factors limit energy expenditure. By contrast, rural areas have been seen as places where heavy work on farms, forestry and mining-related activities leads to high levels of energy expenditure. It was thought that the levels of physical activity in rural areas were much higher than those in cities, and hence that the likelihood of weight gain was much smaller in rural than in urban populations.

Research has shown that in some low-income countries, such as China, people living in urban areas have diets that are distinctly different from those of their rural counterparts[3]. In the past two decades, a shift towards obesogenic diets has promoted weight gain and increased the risk of health

problems related to chronic diseases in urban areas in China[3].

But some research findings have indicated that the levels of overweight and obesity are increasing faster in rural than in urban areas, even in many low- and middle-income countries (LMICs)[4]. This is likely to be linked to the fact that rural areas in LMICs have begun to resemble urban areas, because the modern food supply is now available[5] (Fig. 1) in combination with cheap mechanized devices for farming and transport. Ultra-processed foods are becoming part of the diets of poor people in these countries, and there are reports that infants are even being fed with these foods[6]. Despite these observations, most research and policy efforts have been focused on tackling urbanization as a major driver of obesity, because the general thinking is still that people living in rural areas are much more likely to face hunger and under-nutrition than to be exposed to factors that lead to excessive weight.

All earlier research on BMI trends was based on limited data, and focused on either LMICs or high-income countries[4]. In this context, the paper by the NCD Risk Factor Collaboration is ground-breaking, because it pulls together the latest data from almost all countries to comprehensively examine global BMI trends. The results show that the levels of overweight and obesity are already greater in rural than in urban areas in all high-income countries, and also suggest that the rate of change in many LMICs is such that the levels of overweight and obesity in rural areas will soon match, if not exceed, those in urban areas. Rural hunger, wasting and stunting are rapidly being replaced by overweight and obesity in most regions of the world except sub-Saharan Africa, South Asia and a small number of countries in other areas.

This finding is fundamental, because the