Physics for neuromorphic computing

Danijela Marković, Alice Mizrahi, Damien Querlioz and Julie Grollier

Abstract | Neuromorphic computing takes inspiration from the brain to create energy-efficient hardware for information processing, capable of highly sophisticated tasks. Systems built with standard electronics achieve gains in speed and energy by mimicking the distributed topology of the brain. Scaling-up such systems and improving their energy usage, speed and performance by several orders of magnitude requires a revolution in hardware. We discuss how including more physics in the algorithms and nanoscale materials used for data processing could have a major impact in the field of neuromorphic computing. We review striking results that leverage physics to enhance the computing capabilities of artificial neural networks, using resistive switching materials, photonics, spintronics and other technologies. We discuss the paths that could lead these approaches to maturity, towards low-power, miniaturized chips that could infer and learn in real time.

Brain-inspired algorithms such as artificial neural networks have demonstrated an unprecedented success in machine learning. These algorithms most often run on supercomputers, which, unlike the brain, physically separate core memory and processing units. This slows them down and substantially increases their energy consumption. Neuromorphic computing responds to these problems by taking inspiration from the brain to create energy-efficient hardware for information processing, capable of highly sophisticated tasks.

In this Perspective article, we make the case that building this new hardware necessitates reinventing electronics. We argue that research in physics and material science will be key for creating artificial nanoneurons and synapses, connecting them together in huge numbers, organizing them in complex systems and computing with them efficiently. We describe how some researchers choose to take inspiration from artificial intelligence to move forward in this direction, whereas others prefer taking inspiration from neuroscience, and we highlight recent striking results obtained with these two approaches. Finally, we discuss the challenges and perspectives in neuromorphic physics, which include developing the algorithms and the hardware hand in hand, making substantial advances with small toy systems and building large-scale networks.

The brain as inspiration

Biological brains perform extremely complicated tasks because living beings need to solve complex problems for their survival. The smallest worm analyses data, takes decisions and is able to move towards its goal[1]. Furthermore, brains compute with a remarkably low energy budget because living beings cannot use more energy than they can fetch and store. The human brain categorizes stimuli, predicts outcomes, creates thoughts and ensures vital life functions with a power consumption of only ~20 W. In contrast, present-day digital computers are optimized for high-precision calculations but consume an inordinate amount of energy when they run the type of cognitive tasks that the brain excels at. For example, training a state-of-the-art natural language processing model on a modern supercomputer consumes 1,000 kW h, which is the energy consumed by a human brain to conduct all of its tasks over 6 years (BOX 1 provides further information about this comparison). In addition, the brain is capable of accumulating knowledge over its entire life, and can learn from very few examples. By contrast, current artificial intelligence (AI) algorithms suffer from catastrophic forgetting and must be trained with large comprehensive datasets to accomplish even simple image recognition tasks. Multiple studies suggest that more brain-inspired approaches might suppress these issues[2–6]. Indeed, the brain differs vastly from human-made computing systems, both in terms of topology and in the way it processes information, which explains its different aptitudes.

'Algorithms' of the brain. With its 10^{11} neurons interconnected through 10^{15} synapses, the human brain is a complex system that is dynamic and reconfigurable. It exhibits a wealth of fascinating phenomena for physicists, such as minimization of energy or entropy[7,8], phase transitions and criticality[9], self-oscillations[10,11], chaos[12], synchronization[13,14], stochastic resonance[15] and many more. For many years, physicists have participated in theoretical efforts to understand the algorithms of the brain, contributing to both computer science and computational neuroscience. One of the most striking differences between the brain and computers that use conventional instruction-based algorithms is that the brain can learn from experience. The neural mechanisms that enable learning have been partially elucidated by statistical physics, nonlinear dynamics and complex systems theory. In some cases, learning algorithms rely deeply on physical principles for computing. The most famous examples are Hopfield networks[7] and Boltzmann machines[16], which are derived from Ising spin systems, but many others have been proposed, exploiting nonlinear dynamics effects such as bifurcations and chaos for computing[17–19]. These physics-based approaches mirror nonlinear dynamical approaches that employ spiking neural networks, directly inspired by neurosciences[20,21].

Physical implementation. The brain as a physical substrate of computing is also fundamentally different from general-purpose computers based on digital circuits[11,22]. It is based on biological entities such as synapses and neurons instead of memory blocks and transistors. It leverages the stochasticity of cells to function at very low energy, instead of relying on high-precision circuits as in digital electronics (BOX 2). It uses analogue computation to integrate information inside neuron somas, whereas the communication between neurons is handled by binary spikes, resembling digital communication.

Box 1 | **Comparing the energy efficiency of a brain and a supercomputer**

Here, we compare the brain[152] with a supercomputer training the deep neural network BERT[153,154].

The brain performs and learns multiple tasks in parallel, such as natural language processing, vision, auditory processing, synthesis, social interactions, sensory–motor coordination, basic mathematics and so on. The neural network BERT is state of the art in natural language processing.

Given an energy budget of 1,000 kW h (3.6 GJ), the brain can operate for 6 years. BERT, with the same energy budget, can only be trained during 80 h using 64 T V100 graphics processing units.

The brain consumption of oxygen and glucose represents 20% of the whole body power consumption, amounting to 20 W for an adult. In the supercomputer, the power consumption comes mostly from the central processing unit, the graphics processing units and the dynamic random-access memory.

The brain is composed of about 10^{15} synapses and 10^{11} neurons. The neural network BERT has 10^8 tunable synaptic weights.

It is difficult to establish a fair comparison of the energy consumption of the elementary operations in the brain and in a supercomputer. The energy consumption of spikes and synaptic events in the brain is known, and comes mostly from moving Na^+ and K^+ ions across cell membranes. It amounts to 3.8×10^8 adenosine triphosphate molecules or 60 pJ for a neuron spike, and 1.6×10^5 adenosine triphosphate molecules or 26 fJ for a synaptic event. Such a calculation is more difficult for a supercomputer, which does not contain physical synapses and neurons, and delocalizes the operations in different parts of the processors. What can be said is that training BERT takes 10^6 steps, each composed of 10^5 words and for which all weights are modified. Therefore, the energy per weight modification can be estimated to be 100 pJ.

The figures given above highlight the striking differences in the overall energy consumption of the brain and a supercomputer. It is even more difficult to compare the energy consumption of the brain and a supercomputer for identical tasks. Contrary to artificial neural networks, the brain never stops learning and there is no separation between the costs of inference and learning. Intense task focus on a task in the brain increases the local energy consumption by less than 10%. Being in a coma reduces the consumption by 50% and sleeping by 10–30%. For the supercomputer, we have considered the case in which the neural network is trained on a single task. However, the network architecture (number of layers, neurons and parameters) had already been optimized. The cost of learning from scratch from a new dataset involves tuning hyperparameters and architecture search, which was not included in our calculations, and can lead to up to a 1,000-fold increase in the total power consumption.

Unlike computers, the brain does not feature a unique, perfectly synchronous clock to regulate communication[23]. Some higher-level operations of the brain are hypothesized to embrace collective behaviours emerging in correlated systems for computing, instead of relying on independent circuits with well-defined functions. Finally, the brain exhibits extremely high plasticity due to its billions of adaptive synapses, compared with the limited reconfigurability of digital electronics.

Present-day processors rely on semiconductor physics combined with the Kirchhoff laws of electricity to efficiently implement Boolean logic. But as discussed below, other physical principles and building blocks could be more suitable to implement neuromorphic chips inspired by the brain.

More physics and materials needed

Present-day electronics are not enough. In the brain, neurons, which can be seen as performing computing operations, have a direct access to memory, supported by synapses. Present-day electronics, on the contrary, intrinsically separate memory and computing into distinct physical units, between which data must be carried back and forth. This 'von Neumann bottleneck' is an issue for AI algorithms that require reading considerable amounts of data at each step, performing complicated operations on this data and then writing the results back to memory[24]. This data transfer process slows down computing and considerably increases the energy consumption for learning and inference.

In light of this problem, the general paradigm in neuromorphic computing is to take inspiration from the topology of the brain to build circuits composed of physical neurons interconnected by physical synapses that implement memory in situ, in a non-volatile way, thus dramatically cutting the need to move data around the circuit and allowing considerable gains in speed and energy efficiency. However, this architecture is unfortunately complicated to implement if CMOS technology alone is used. Dozens of transistors are needed to emulate each neuron, and additional external memories are required to implement synapses. CMOS-based artificial neurons and synapses are typically several micrometres wide[25] despite the size of transistors being tens of nanometres, and the number of physical neurons and synapses that can be integrated in a CMOS chip is inherently limited by the chip area. This loss of density is a problem because neural networks require many neurons and synapses to perform useful tasks: typical image recognition algorithms today comprise millions of neurons and synapses on average. Large numbers of neurons and synapses can be obtained by assembling chips together, but the whole system becomes bulky (the most complex versions of existing neuromorphic systems such as Spinnaker or Brainscales[26] can occupy several cubic metres), and much of the energy efficiency is lost in the interconnects between chips. Nanodevices that can mimic important features of neurons and synapses at the nanoscale — features such as nonlinearity, memory and the ability to learn — are required if it is to be possible to build low-power chips comprising several millions of neurons and synapses.

Finally, another drawback of CMOS technology is that using CMOS devices alone makes it difficult to achieve a high degree of interconnection between neurons. The brain features an average of 10,000 synapses per neuron. Such connectivity is impossible to reproduce with CMOS technology, which is usually confined to two dimensions, with limited fan-out. It is also difficult to efficiently and evenly supply energy to components in the circuit. By contrast, the brain is 3D, axons and dendrites provide high fan-in and fan-out, and blood efficiently distributes energy to the whole system.

The requirements of neuromorphic computing. The above considerations highlight some essential needs of neuromorphic computing that could be addressed by new materials and physical phenomena. A goal of research is to develop nanoscale devices that imitate neurons and synapses with low energy consumption and high endurance, that are easy to address (read and write) in large networks, that provide signal gain and memory, that are tunable, active, dynamical, reconfigurable and multifunctional, that provide large fan-in and fan-out, large interconnectivity, can be self-assembled, can form 3D interconnects and are easily manufacturable in large quantities and, of course, at low cost.

AI and neuroscience as guides

How can these general features of the brain be transformed in important contributions to the field of neuromorphic computing? As there is not yet a precise model of how the brain works, two main approaches

are currently in development. The first approach is to map conventional neural network algorithms — that are currently used in AI — to dedicated physical systems, to improve the energy efficiency of their implementation. The second approach is to reach beyond such algorithms, by taking inspiration from neuroscience to equip artificial neural networks with additional features and dynamics, in the hope of achieving more complex computing.

Mapping AI algorithms to physical systems. Since 2012, considerable progress has been made in developing AI algorithms based on algorithms known as artificial neural networks[27]. These algorithms can now beat humans at pattern recognition tasks and at sophisticated games such as Poker or Go[28]. However, the hardware on which they run, such as graphical processing units or tensor processing units, limits their development outside large and particularly energy-intensive data centres[29]. Therefore, developing hardware that is better suited to run current neural networks is an important challenge.

The neural networks that are best able to address real applications today are called deep networks because they are composed of a large number of neuron layers, ranging from four to hundreds (FIG. 1a). They typically take as input high-dimensional data, such as megapixel images, and reduce the dimension by outputting only a few classes, designating, for example, the content of the image (for instance, cat or dog). This can be achieved using simple formal neurons and synapses, which retain only basic features of their biological counterparts. First, in most artificial neural networks, neurons are reduced to a mathematical function applied to their real-valued inputs. This activation function should be nonlinear, such as a sigmoid or a rectified-linear unit, so that the neural network can approximate generic nonlinear functions. Second, synapses are valves for information, configuring how it flows in the network, guiding input information to the relevant output. These valves can transmit information positively as well as negatively and are described by real-valued synaptic weights. The input to a neuron in layer $k+1$ is the output of neurons in layer k multiplied by synaptic weights. This weighted sum operation, also called multiply and accumulate, is a crucial computation of artificial neural networks.

The weights are initially chosen randomly, and the network needs to learn the weights that allow it to perform its task. The most successful known way to train networks is through supervised learning. In supervised learning, labelled examples are presented to the network. Its output is then compared with the known desired output and an error is numerically computed. This error is then backpropagated through the network to change each weight in proportion to its contribution to the error. After several batches of examples, the network can eventually converge until the error is minimal, and it can generalize when presented with examples that it has never seen during training.

Neuromorphic chips have the potential to accelerate both the inference phase (in which a network is presented with an input and computes the output) and the training phase. A challenge is to build systems compatible with the hierarchical layered structure of the most powerful neural networks today, in which neuron outputs of one layer naturally feed the synaptic inputs of the next layer. So far, there are two principal technologies for which potentially scalable architectures have been proposed for mapping deep networks on-chip using physics: hybrid CMOS/memristive systems (FIG. 1b) and photonic systems (FIG. 1c).

In hybrid CMOS/memristive systems, neurons are made of analogue or digital CMOS, and the information that flows through the network is electrical current. To regulate the information flow, a synapse should therefore act as a valve for the current. Such a valve can be implemented with a memristor (short for 'memory–resistor', and also known as a resistive switching device). These devices are resistors with non-volatile analogue conductance states that are tunable by an applied voltage (FIG. 2). The dimensions of a memristor can be reduced to the nanoscale. Such features can be obtained through different physical effects within a wide range of materials (FIG. 2a–g). Effects and materials are extremely diverse and will be discussed in more details further. The most widely studied are: electric-field-induced creation and control of nanoscale conductive filaments bridging two metallic electrodes separated by an insulating oxide such as tantalum or hafnium oxide[30] (FIG. 2a); phase transitions (FIG. 2b) leading to conductance changes in materials such as chalcogenides (structural phase transition, caused by the Joule effect)[31] or strongly correlated electron systems[32] (electronic phase transition, caused by electric field); and voltage-induced control of the ferroelectric configuration in the insulating barrier of ferroelectric tunnel junctions[33] (FIG. 2c).

When these memristors are arranged in a crossbar array configuration, they can be used to fully connect a layer of neurons to the next one[34,35] (FIG. 1b). The current going out of each of the electrodes at the bottom of the array is indeed the sum of the input voltages (applied at each row of the array), each weighted by the conductance of the memristor in the column. Therefore, memristors directly implement the multiply-and-accumulate operation through Kirchhoff's laws (two memristors are necessary for one signed weight), an operation that for CMOS devices is costly in terms of silicon area and energy consumption. Purely resistive arrays have a limited size due to current sneak paths (that is, current flowing from a row to a column through several memristors instead of one). This maximum size can reach a few hundred by a few hundred memristors, for devices with nonlinear current–voltage characteristics that inherently limit the impact of sneak paths[36]. One road currently explored to achieve higher connectivity consists of assembling different crossbars together through CMOS-based buffers, either side by side[37] or on top of each other[38]. Another route is to add a selector device below each memristor, that is, a device controlling whether current can flow through or not, which can be either a transistor[39–41] or a volatile memristive switch[42–44]. The advantages of memristors in terms of compacity and energy consumption might not be limited to inference. It has been recently estimated through combined experiments and simulations that memristive systems could enable training neural networks with a 100-fold gain in

Box 2 | **Stochasticity in the brain**

In the brain, ion channels open and close stochastically. The resulting current is a random telegraph signal. As there are typically 100 ion channels open simultaneously, this overall noise level in the current can be estimated to $\sqrt{100}$ / 100 = 10% fluctuations[155]. This degree of fluctuation is considerable compared with the reliability of conventional microelectronic circuits.

In addition, synaptic release is probabilistic[156]. The release probability is highly synapse dependent and varies with parameters such as the spiking rate and temperature, with typical values around 25%. This mechanism differs fundamentally from the deterministic switching of CMOS gates.

An interpretation of this stochasticity is that the brain may maximize not the information transfer itself but rather the ratio between information transfer and energy[60,157].

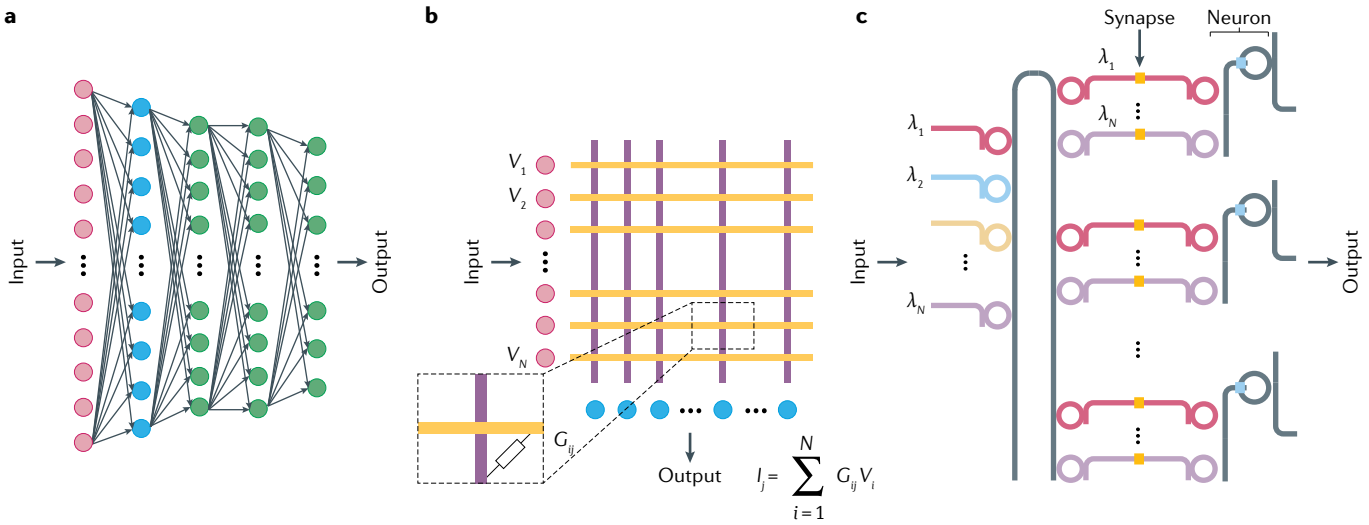

Fig. 1 | **Hardware for deep neural networks. a** | Deep neural network composed of layers of neurons (circles) connected by synapses (arrows), performing classification of its inputs. **b** | Memristive crossbar array connecting the first two layers of *N* neurons[34]. Such crossbars can be stacked to add the following layers of neurons. The inset represents a single memristor cell, vertically connecting a row and a column. The pre-synaptic CMOS neurons (red) apply voltages to the rows. The output current I_j at each column *j* is the sum of all input voltages V_i from rows *i*, weighted by the memristor conductances G_{ij}. An amplifier at each column drives the post-synaptic CMOS neuron (blue). **c** | Optical neural network composed of circular resonators that couple inputs of different wavelengths λ_i (in different colours) to a neuron[46] (in grey). Synapses (orange squares) and neural activation functions (green squares) are implemented by phase-change materials.

terms of energy consumption and speed compared with graphical processing cards[45]. Nevertheless, achieving learning with memristors raises challenges, which are addressed in the final section of this Perspective article.

Another neuromorphic computing approach that is scalable to deep networks is based on photonic neural networks. These networks can be made with optical components only or mix optics with electronics using optoelectronic devices. The nonlinearity needed for implementing neurons can be provided by the Kerr effect or by carrier heating in semiconducting lasers. The synaptic multiply-and-accumulate function can be implemented by combining multiple interferometers or by modifying the transmission of optical waveguides with optically active phase-change materials deposited on top[46–48] (FIG. 1c). Simple tasks have been demonstrated with these systems, such as vowel recognition[47]. The advantages of using optics for computing are the possibilities of conveying large amounts of information in parallel on a single fibre[49] or waveguide through massive wavelength multiplexing, and of building purely passive neural networks with an extremely low energy consumption[50]. Furthermore, devices that explore 3D photonic interconnects and brain-inspired topologies are being developed[51]. However, the size of neurons and synapses is ultimately limited by the optical wavelength and might be difficult to shrink below a few micrometres. Methods of size reduction that have been investigated include fabrication optimization[52] and the use of plasmonics[53]. Furthermore, reaching nonlinear optical regimes requires high input power, resulting in a considerable energy cost, on top of the cost of converting the information to light. Low-energy nonlinearities are thus being investigated[54].

Matching neuroscience-inspired concepts to hardware. The brain is far more complex than present-day AI algorithms. Biological neurons cannot be reduced to a nonlinear function between their input and output, and operate with spikes (also called action potentials). They are thought to provide the brain with energy efficiency and increased functionality, as information can be encoded both in their rate and in the timing between them. Important neuromorphic hardware research is, therefore, focusing on implementing neural networks that spike. For example, the deep neural networks of AI can be transferred to spiking ones, with minimal or no accuracy degradation[55,56]. The spiking implementations of such networks, in which the neurons with little spiking activity consume little energy, may bring important energy savings[57]. However, such designs might miss opportunities offered by taking inspiration from neuroscience. Biological neurons possess many other features than their spikes. They are leaky integrators, they feature memory, can be stochastic, and can oscillate and synchronize (FIG. 3a). They are spatially extended, with different functional compartments integrating signals coming from different areas[58,59]. Biological synapses are also more than analogue weights. They are leaky memories, and they have different timescales and different state parameters ruling their modifications. They can also be highly stochastic, as some synapses transmit only a fraction of the spikes that they receive[60] (BOX 2). Frequently ignored brain components also play important roles and might be beneficial to emulate: dendrites seem able to deploy extremely complex computations[61,62], and astrocytes (FIG. 3a) seem to play a critical role in neuron regulation[63]. All these properties can bring additional features to artificial neural networks and are therefore interesting to implement and test.

These ideas are especially promising if they can be implemented at low energy through the intrinsic properties of materials and related physical effects. This field of research was originally focused on exploiting the fact that the leakage current of a transistor has exponential dependence on voltage[64–66]. Since the late 2000s, a wide range of other physical phenomena (FIG. 2) have been used to mimic interesting properties of synapses and neurons.

In particular, oxide electronics can imitate the multifunctionality of synapses and neurons. Biological synapses memorize meaningful information for long time periods but rapidly forget unimportant data. Conductive bridge devices (FIG. 2a) not only

act as current valves, as discussed above, but also can emulate this dual long- and short-term memory nature of synapses[67,68]. Low-amplitude voltage pulses applied to the device trigger metallic filament growth between the electrodes. If the pulses are infrequent, the filament shrinks back, giving rise to short-term, leaky memory. However, if the pulses are frequently repeated, the filament does not have time to relax and grows until it strongly bridges the two electrodes, giving rise to long-term memory (FIG. 2h). Another feature of biological neurons that can be imitated is the way they emit spikes when their membrane potential exceeds a threshold. Materials exhibiting electronic phase transitions, such as Mott insulators, can emulate spiking neurons. When they are excited with series of incoming voltage spikes, an avalanche effect occurs and locally triggers the transition to a conductive state. If the input remains silent, the material slowly relaxes back to its original state. These two phases successively mimic neuron depolarization and repolarization during an action potential[69,70]. Volatile resistive switching devices exhibiting thermally induced negative differential resistance can also be combined with additional capacitors, resistors or transistors to generate trains of spikes under a constant input voltage[69,71–74] (FIG. 2i). The spiking behaviour in these systems is rich, including periodic spiking, chaotic spiking and bursting, as in biological neurons[72,75,76]. Finally, in the brain, neurons influence each other's dynamics through their coupling via synapses, and can synchronize if synapses are strong[13]. The synchronization of neuron spikes enhances their efficiency for modifying synapses and triggers long-term learning[77]. Memristive oxide synapses can control the dynamical coupling between spiking CMOS neurons, and lead to their synchronization[78].

Chalcogenide-based phase-change memories (FIG. 2b) are another important lead for providing devices for neuromorphic computing, and stand out with regards to memristive oxides due to their technological maturity[79]. They also exhibit memristive behaviour and can implement non-volatile synapses not only for electronics[45,80,81] but also for optics: they can be used as a coating on optical waveguides to tune the transmitted light[46]. The stochasticity of the phase transition between the amorphous and crystalline states has also been harnessed to implement stochastic nanoneurons whose membrane potential is represented by the phase configuration of the device[82].

Ferroelectric tunnel junctions (FIG. 2c) and ferroelectric field-effect transistors exploit polarization switching in ferroelectric materials to induce variations of several orders of magnitude in output signals. Multidomain polarization reversal through domain wall nucleation and propagation has been leveraged for the implementation of memristors and learning[33,83]. The advantage of ferroelectric devices is the purely electronic nature of ferroelectric-based resistive switching, which promises high endurance, associated with large output signal dynamic range. A major challenge is to integrate ferroelectric devices on silicon. Recent results indicate that HfO_2 is a material that holds promise in this direction[84].

Spintronic devices (FIG. 2d) can also provide building blocks for neuromorphic

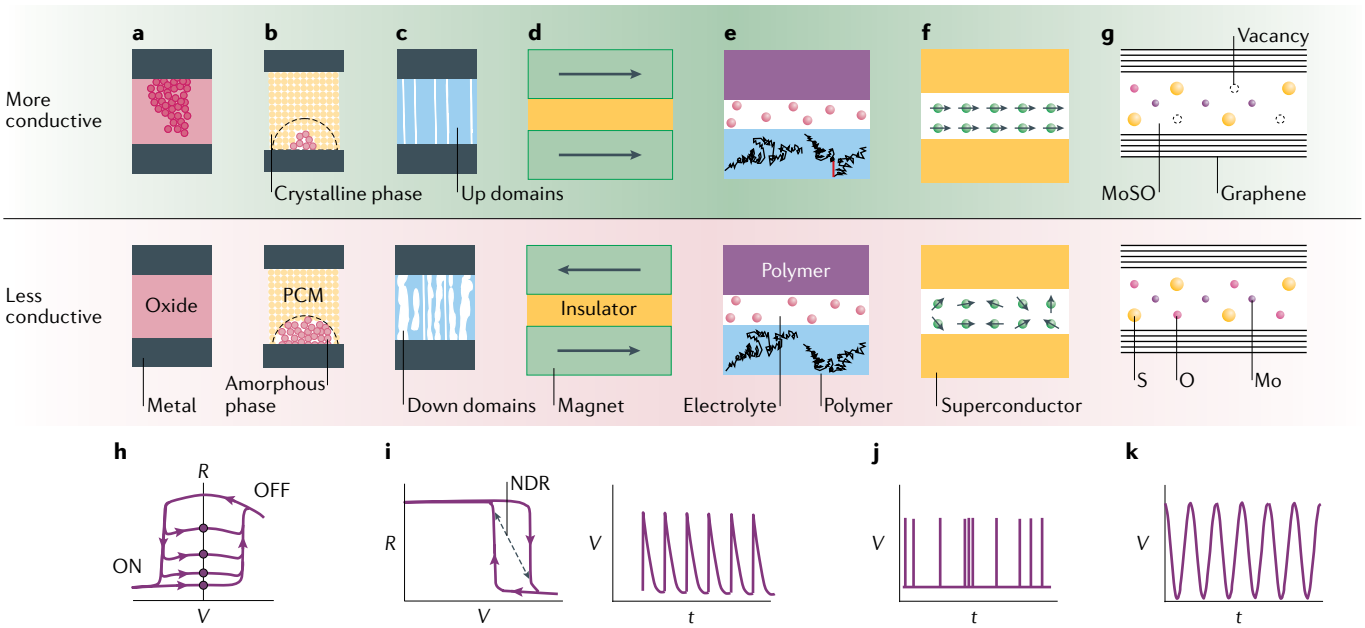

Fig. 2 | **Materials and physics for neuromorphic computing. a–g** | Different physical implementations of synapses. The top row depicts the high-conductivity synaptic states, and the bottom row depicts the low-conductivity synaptic states. In a filamentary device (oxide or conductive bridge; part **a**), the size of the filament between the top and bottom electrodes determines the resistance. In chalcogenide-based phase-change memory (PCM; part **b**), the size of the amorphous region determines the resistance. In a ferroelectric tunnel junction (part **c**), the resistance is given by the fraction of ferroelectric domains pointing downwards. In a magnetic tunnel junction (part **d**), the relative orientation of the magnetic layers determines the resistance. In an organic electrochemical device[90] (part **e**), conduction is assured by the positive ions in the electrolyte layer. A positive pre-synaptic voltage causes polymer reduction in the post-synaptic electrode, which absorbs some of the conducting ions and thus decreases the synaptic conductance. In a Josephson junction with a magnetic barrier[96] (part **f**), the magnetic order in the barrier determines the resistance. In a graphene/MoSO/graphene van der Waals heterostructure[97] (part **g**), the concentration of oxygen vacancies determines the resistance. **h** | Electrical resistance *R* versus applied voltage *V* of a non-volatile memristor, exhibiting multiple resistance states. **i** | Electrical resistance versus applied voltage of a volatile memristive switch with negative differential resistance (NDR), and voltage spikes (as a function of time *t*) generated by such a device. **j** | Voltage spikes generated by a resistively and capacitively shunted Josephson junction. **k** | Voltage versus time across a magnetic tunnel junction with current-induced sustained oscillations.

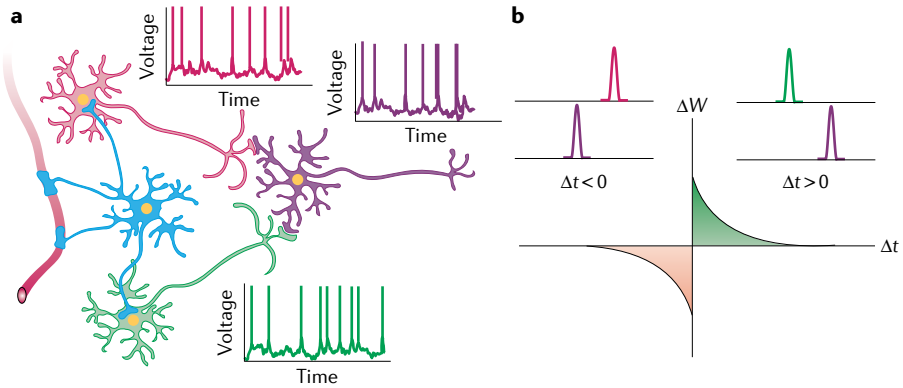

Fig. 3 | **Biologically inspired neuromorphic computing. a** | Two pre-synaptic neurons (red and green) connected to a post-synaptic neuron (purple). An astrocyte (blue) provides nutriments and energy from a capillary to neuron somas. All the neurons emit spike trains of action potential. **b** | Spike-time-dependent plasticity. The pre-synaptic spike from the green neuron arrives shortly before the post-synaptic spike from the purple neuron ($\Delta t > 0$), suggesting a correlation between these two events. The synapse connecting the two neurons is meaningful and its weight is increased ($\Delta W > 0$). On the contrary, the pre-synaptic spike from the red neuron arrives shortly after the post-synaptic spike. The weight of the synapse that connects them is decreased ($\Delta W < 0$).

computing based on materials and processes with high technological maturity. Their distinctive feature is their outstanding cyclability[85] and the availability of accurate models for their behaviour[86]. They also provide a multifunctionality capability, particularly promising for neuromorphic computing, as devices can be tuned, for example, between an artificial synapse or a neuron by careful sizing[87]. Their stochastic switching and rich nonlinear dynamics have been exploited to demonstrate different types of nanoneuron that exploit magnetization dynamics for computing[88,89] (FIG. 2k).

Another emerging lead in neuromorphic research is the use of fully organic materials, some of which may have the unique advantages of biocompatibility, and possibly low-cost fabrication through the use of printing technologies. Organic memristive devices are mechanically flexible, multifunctional and often low power. Their application to neuromorphic computing includes harnessing their electrochemistry to implement low-voltage artificial synapses[90] and neurons[91] (FIG. 2e). However, these devices tend to suffer from slow operation due to the low mobility of carriers in organic materials, high levels of variability, lack of reliability due to the instability of organic materials, and poor retention.

Memory technologies are not the only lead for physics-based neuromorphic computing. Nanoelectromechanical systems provide another type of dynamical system that can emulate neurons through their resonant or their self-sustained oscillations[92]. Of distinct interest for neuromorphic computing is their oscillation frequency, which is compatible with audible sounds, opening the route to on-chip voice recognition[93].

More unexpectedly, flux quantization in superconductive Josephson junctions can also provide an analogue to voltage spikes (FIG. 2j). Rapid single flux quantum circuits can be adapted to emulate neuron spike emission and interconnection, with the advantage of extremely low power dissipation[94,95]. Superconductive synapses have also been experimentally demonstrated. For instance, magnetic cluster insertion in the junction barrier can transform a Josephson junction in an analogue non-volatile valve for the input current[96] (FIG. 2f). Josephson junctions may bring neuromorphic computing to cryogenic electronics, and open highly original applications, such as for basic or medical physics experiments.

At an even more prospective level, 2D materials, which are composed of atomically thin layers, can be gated and assembled in van der Waals heterostructures to generate a variety of behaviours, such as memristive behaviour through ionic motion[97,98] (FIG. 2g). Resistive switching in 1D and 2D materials can be controlled using light, an ability that provides the possibility to build bioinspired optical sensors for neuromorphic chips. This control by light could possibly lead to richer behaviours than those seen in memristive oxides and phase-change memories.

Finally, optical devices provide another venue for physics-based neuromorphic computing, offering the potential for very high-speed operation. In particular, neuromorphic computing can harness optical nonlinear dynamical phenomena to implement dynamical artificial neural networks. Biological neural networks are dynamic and rely on recurrent loops, and these features are useful to implement short-term memory — the memory the brain uses to remember sequences of numbers or the beginning of a sentence[99]. The low attenuation of optical fibres makes it possible to build connections between different points of a system and has been leveraged to build recurrence loops. Delayed feedback optical oscillators can implement short-term working memory[100]. Using an approach called reservoir computing, they can even perform classification and prediction tasks that require a dynamically stored memory[101–103]. Spatial light modulators are currently used to increase the number of neurons or binary synapses in optical networks[104].

Among these different bioinspired approaches, the most interesting are probably those that will allow crafting of synapses and neurons using the same technology, thus allowing easier co-integration within large-scale systems and implementation of learning algorithms through the physics of these devices.

Building neuromorphic systems in a lab

Most of the demonstrations discussed above concern single devices, or at most tens of devices. In this section, we provide guidelines on how to build larger-scale neuromorphic systems. We describe toy systems and explain what needs to be done to move from individual-device physics to systems.

Algorithms and physical computing substrates should be developed hand in hand. Neuromorphic computing requires physics and materials science to engineer devices with remarkable functions, and bioinspired computing models to develop algorithms. These two avenues need to be thought about and developed together. This need holds equally for both AI and neuroscience-inspired computing.

For example, there are two main challenges at the device level regarding the implementation of learning-capable neuromorphic systems. The first is precision. Learning through standard backpropagation, the most effective procedure for training deep AI networks, involves updating the weights after each batch of examples is presented, with variations that can be as small as 0.01% of the value of the weight itself, owing to vanishing gradient effects[105].

These subtle adjustments are not an issue when weights are encoded in 32-bit floating point, as is usually done on graphics cards. However, most nanodevices are inherently noisy and have effective weight values that do not vary perfectly continuously[106], making small weight adjustments difficult to achieve. The second challenge is to perform weight-independent weight variations with an applied voltage or current. Device nonlinearities, which are often at the core of memristor or any artificial nanosynapse physics, will indeed typically prevent convergence of backpropagation[45,106]. There is therefore today a considerable effort to produce devices with smooth and linear features. That said, the backpropagation learning rule intrinsically copes well with device-to-device variability, as discussed in the final section of this Perspective article.

In parallel to research aimed at perfecting present-day devices, a considerable effort consists of modifying existing algorithms so that they can work despite the imperfections and unreliability of the physical substrate[37,81,107]. Device–algorithm co-design, such as adaptive quantization, makes it possible to adapt algorithms to the particular limitations of devices, such as the finite number of conductance states in the memristors that can only implement a certain number of discrete weight values[108]. Success has been obtained in deep networks that have binary weights instead of analogue ones, a modification that constitutes a major simplification for hardware implementations[41,109], and which can achieve high accuracy on image recognition[110] and signal processing[111] tasks. State-of-the-art results have been obtained during the inference phase; however, online learning still requires real values for the weights[112–114]. In parallel, pruning is a method in machine learning that removes unimportant weights and thus spares the computing power. It has recently been adapted for neuromorphic networks as soft-pruning, in which values of the weak weights are rounded during the training phase, thus reducing the energy consumption while preserving accuracy[115,116]. Physics-inspired restricted Boltzmann machine algorithms are another interesting lead, as they are more tolerant to memristor nonlinearity during learning than more traditional neural networks are[117,118]. Other solutions work both at the algorithm and device level. For example, the stochasticity and nonlinearity of memristive synapses can be alleviated by combining several devices together[81], or by using a linear capacitor as a weight for learning, but regularly transferring the weight value to a non-volatile memristor before it is forgotten[45].

That said, the brain is able to learn with unreliable and stochastic components. It should therefore be possible to invent novel algorithms that allow learning with elementary nanodevices that are imperfect, dynamical and noisy. In this direction, unsupervised learning is particularly interesting. Such networks, owing to their lack of supervision, are extremely unconstrained, and neurons are free concerning which feature they specialize in learning. This adaptability tends to make unsupervised learning highly resilient to device imperfections[119,120]. The natural fitness of nanodevices for unsupervised learning is appealing, as one of the main challenges in AI today is to develop algorithms that do not need millions of labelled examples to learn from. A promising unsupervised learning algorithm is spike-timing-dependent plasticity (STDP)[121], a learning rule inspired by measurements in neuroscience[122]. In this rule, the synaptic weight of synapses is modified in a way that depends only on the timing of spikes occurring on both sides of a synapse (FIG. 3b). Unlike error backpropagation, this learning strategy is spatially local and memristive devices can implement it naturally. A first strategy is to connect non-volatile memristors with neurons that exhibit pulses with a specifically designed shape with positive then negative parts[83,123]. Overlap occurring between voltage pulses on both sides of the synapse can cause the application of high voltages across a device and the modification of its conductance, therefore changing the synaptic weight. Another elegant strategy is to harness the volatility of some memristive devices to implement the learning rule directly through device physics[124,125]. For example, a spike on one side of the synapse can briefly elevate the temperature of a device for a short period through the Joule effect, causing any subsequent spike during this period to have an enhanced effect on device conductance. Pattern recognition through STDP has been demonstrated with small ensembles of memristors[126,127], and with larger ensembles of phase-change memories[118]. At present, the biggest challenge for scaling STDP-based systems might not come from devices and materials, but from the limitations of the STDP rule itself. Newer theories are needed to extend the cognitive capabilities of STDP to complex multilayer systems, and the availability of memristive devices is one of the core motivations for such research[128,129].

One can even imagine algorithms that exploit nanodevice defects or dynamics for information processing. For example, noise can be harnessed for computing in many different ways, including stochastic resonance[130], exploration of energy minima[89,131] and diffusion[132]. Furthermore, brain networks exhibit criticality, phase transitions and synchronization — can these physical effects be harnessed in materio for computing and learning? These questions are currently explored by many groups, theoretically and experimentally.

Small-scale 'toy' systems. Building large-scale systems often requires access to a CMOS foundry and to industrial partners. However, academic laboratories can make important contributions to system-level developments by developing 'toy' physical neuromorphic systems. The idea is the same as that of the well-known 'toy models' in physics: develop small physical neuromorphic systems to test some hypothesis and draw some conclusions. Recent experiments by several groups have shown that realizing such systems can bring important insights that would not necessarily be identified by theoretical investigations or single-device studies.

In the following, we illustrate this approach by describing our own work, focused on the development of experimental toy neural networks composed of spintronic nano-oscillators (FIG. 2d,k) as neurons. In our first experiment[88], we wanted to check whether a spintronic oscillator could be used as a neuron to compute. For this purpose, we used the approach of reservoir computing, which makes it possible to emulate a full neural network with a single oscillator multiplexed in time[101] (FIG. 4a). We learned from this experiment that oscillator noise is a severe issue for reservoir computing — an issue that will also apply to other forms of neural networks. In such a time-multiplexed approach, the oscillator output is sampled, and the resulting data points are linearly combined to form the output. To obtain reliable pattern recognition, the oscillator should therefore not drift, and it should feature a high signal-to-noise ratio. This is problematic at the nanoscale, as fluctuations and drift effects often arise due to the small volume of devices[87]. However, we found an exploitable range of conditions in which spintronic nano-oscillators have enough stability and signal-to-noise ratio to obtain performance at the state of the art of macroscopic oscillators and even software[88]. This is due to the exceptional cyclability of spintronic nanodevices and the possibility

to finely tune their dynamical response through current and fields.

Computing with noisier oscillators is possible but requires engineering redundancy at the neuron level so that the ensemble average can suppress the effects of noise. We tested this strategy with another kind of oscillator, namely superparamagnetic tunnel junctions[89,130], which are noisy to the point that they behave in an entirely stochastic fashion. We took inspiration from a neuroscience theory called population coding to implement the computation of various nonlinear functions[133].

After our first experiment with a single spintronic nano-oscillator, we moved on to our next neuromorphic system: a small neural network formed of four coupled spintronic nano-oscillators[134] (FIG. 4b). We investigated whether it was possible to achieve pattern recognition by synchronizing some of the oscillators to incoming alternating inputs, which encoded the signal to classify in their frequency. The experiment worked and was able to classify spoken vowels[134]. However, it required the oscillators to be highly tunable in frequency. Nano-oscillators indeed feature an unavoidable variability in their properties. This variability in the basic components of a neural network can be compensated during training, which was achieved, in our small neural network, by exploiting the high tunability of the frequency of spintronic oscillators with the injected direct current (reaching tens of percent of the base frequency).

Developing large-scale systems. The ultimate goal of neuromorphic computing development is to build large-scale systems comprising millions of components. Therefore, care should be taken to study materials and physics that are scalable. A first concern is that, so far, most neuromorphic research groups have focused on developing either large-scale memristor arrays to be used as synapses or biologically plausible neuron devices. The integration of energy-efficient nanoscale synaptic and neuron devices is timely to realize large-scale neuromorphic systems[135]. More fundamentally, scaling neuromorphic systems relying on complex physical phenomena raises several major challenges. First, densely and efficiently interconnecting nanosynapses and nanoneurons is a substantial difficulty in neuromorphic computing. There are two strategies to tackle this problem. It is possible to step away from bioinspiration and use CMOS circuits for interconnection. In that case, conventional electrical engineering methods can be used to route signals between neurons and synapses, in particular, multiplexing techniques and routers to limit the number of interconnections. Then, the difficulty of achieving high-density connection in electronics can be compensated by the availability of fast digital circuits to achieve routing. The danger in this situation is losing the advantages of using emerging nanotechnology in the first place, due to the CMOS overhead in terms of area and energy consumption. For this purpose, the CMOS circuits need to optimally exploit the physical properties of the nanotechnology, through a radical co-design approach. The other strategy to achieve dense and efficient interconnections is to harness physics, materials science and nanotechnology to densely connect the neural network at the system level, and attain densities above 1,000 synapses per neuron, approaching the level of interconnection of the brain. A pioneering idea in this direction is that of 3D crossnets[38,136], in which memristor crossbar arrays are stacked on top of each other. Other approaches include wireless communication between neurons and synapses, using optics[101,104] and microwaves[134], and self-assembly of nanoparticles or nanowires in 2D and 3D[137].

The second challenge for realizing large-scale neuromorphic systems with physics is to deal with nanodevice variability. Training the network on-chip is the best way to compensate for these device variations. Indeed, through learning, each component (synapse and, in some cases, neuron) can be tuned to deliver the desired output, taking into account the specificities of the devices it is connected to[45,118,131]. Another exciting lead is to implement systems in which the device variability is seen as an asset, not an issue. For example, the CMOS-based Braindrop system[138] fully embraces device mismatch, as transistor variability is used to provide a collection of circuits with a wide range of behaviours, which provide basis functions for building complex functionalities. Alternatively, the variability of memristive devices can be leveraged to implement algorithms such as Markov chain Monte Carlo that require a large quantity of random numbers, which the nanodevices naturally provide through their cycle-to-cycle variability[131].

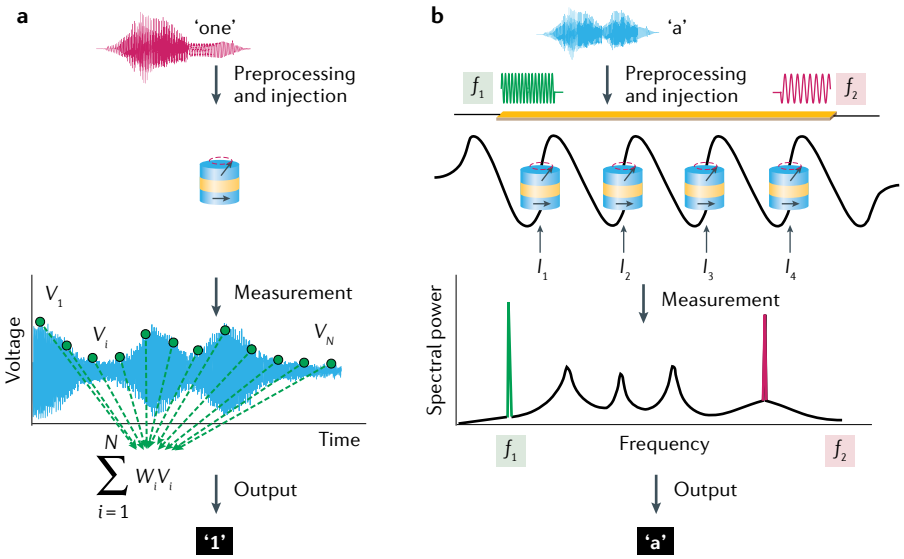

Fig. 4 | **Toy neuromorphic systems with spintronics. a** | Reservoir computing with a single time-multiplexed spintronic nano-oscillator. The spoken digit to recognize (here, 'one') is preprocessed into an input injected to the nano-oscillator that emulates different neurons ($i = 1$ to N) at different time steps. The output is a linear combination of the nano-oscillator emitted voltage V_i with weights W_i at different time steps. **b** | Pattern recognition from synchronization patterns of four coupled spintronic nano-oscillators. The spoken vowel to recognize is preprocessed into two input RF signals f_1 and f_2 injected to the network by a common waveguide. The nano-oscillators are electrically coupled to each other and capable of synchronizing to the input signals. Their frequencies are tuned by injection of individual d.c. currents $I_1 \ldots I_4$. Each synchronization pattern (here, oscillator 4 is synchronized to input frequency f_2) corresponds to one learned vowel (here, 'a').

Outlook

In this Perspective article, we have highlighted the relevance of physics for neuromorphic computing. Neuromorphic computing is a rapidly advancing field, in which many approaches are currently tested, and which all have their advantages and disadvantages (TABLE 1). The question of

Table 1 | **Comparison of the different approaches to neuromorphic computing**

Technology	CMOS synapses and neurons	Resistive switching synapses with CMOS neurons	Photonic synapses and neurons	Spintronic synapses and neurons	Superconductive synapses and neurons
Connections	Wires	Wires	Light	Microwaves	Wires or microwaves
Minimum lateral size of neuron	10 μm	10 μm	100 μm	10 nm	20 nm
Minimum lateral size of synapse	10 μm	10 nm	1 μm	10 nm	20 nm
Advantages	Commercially available	Nanoscale synapse, technology ready	Wavelength multiplexing, can be completely passive (low energy consumption[150,151])	Nanoscale synapses and neurons, almost commercial technology	Low energy consumption beside cryogenic requirements, all identical spikes
Disadvantages	Size of neurons and synapses, no in-memory computing	Size of neurons, complex wiring	Size of neurons and synapses, dissipation required for nonlinearity	Scalability yet to be demonstrated	Scalability yet to be demonstrated
Chip capabilities	Inference and learning	Inference	No chip	No chip	No chip

what materials and physical principles will be used in future neural network hardware is still open-ended.

Qubits and quantum oscillators are conspicuously absent in TABLE 1. Until recently, the fields of quantum computing and neuromorphic computing were evolving in parallel. However, cross-fertilization between the two domains has started and is likely to bring remarkable results. Building physical neural networks of quantum neurons and synapses opens possibilities to use quantum superposition and entanglement to process information in parallel, and to use high-dimensional state space to implement deep neural networks, while taking advantage of neuromorphic computing to deal with noise and device variations[139–141].

As we have pointed out, neuromorphic computing with emerging nanodevices, physics and materials is a fast-evolving field. Recently, different systems using more than a million memristors have been reported[142,143]. These first fully integrated systems are functional and report reasonable raw accuracy on practical tasks, nevertheless remaining much lower, by several percent in the success rate, than the standards of software AI. However, it should be understood that due to the long delays associated with fabricating and testing functional electronic systems incorporating emerging technologies, these chips were designed 1 to 2 years before the systems were demonstrated, and therefore do not incorporate the latest progresses of the field. We should expect exciting new results to come from both academic and industry labs soon. At present, one of the most optimized fully integrated inference systems reported[144] achieves 94.4% accuracy on the canonical Modified National Institute of Standards and Technology database (MNIST) task of handwritten digit recognition. By contrast, to this day, achieving large functional systems capable of learning remains a challenge. Most works use hybrid approaches in which a part of the system is simulated in software[45,131]. Nevertheless, a fully hardware system implementing STDP learning with 1.4 million synapses has been demonstrated[143]. Despite some limitations inherent to early-generation developments, it shows the way for future brain-inspired hardware exploiting physics.

Substantial progress is expected in upcoming years, with the increased participation of industrial laboratories. As the leading semiconductor manufacturers have now integrated emerging memory technologies (resistive[145], magneto-electronic[146] and phase-change[79] memories) into advanced CMOS processes, there is high hope for large-scale — and

Glossary

Astrocytes
Cells in the brain that assist neurons for blood and metabolism regulation. Evidence points to their role in communication and processing.

Axons
Nerve fibres that conduct the action potentials away from the soma to other neurons.

Boltzmann machines
Stochastic Hopfield networks that use a Boltzmann distribution inspired by statistical physics in their sampling function.

Dendrites
Branched extensions of neurons, which conduct stimulation received from another neuron towards the neuron soma.

Fan-out
Typical number of connections spreading from a given point in a circuit. In the brain, one neuron is connected to 10,000 others, that is, it has a 10,000 fan-out.

Hopfield networks
Specific type of recurrent neural network (neural network containing recurrent loops) that has neurons functioning as binary threshold nodes.

Kerr effect
Change of the refractive index of a material due to an applied electric field, proportional to the square of the field amplitude.

Modified National Institute of Standards and Technology database
(MNIST). Dataset of 28×28 pixel images of handwritten digits, widely used as a benchmark for image classification.

Recurrent loops
Connections from neurons to themselves or to neurons in preceding layers (that is, on the input side) of the network. These loops are key for processing time-varying inputs.

Reservoir computing
Specific type of neural network for which an assembly of neurons — the reservoir — has fixed random recurrent connections, and only connections from the reservoir to the output are trained.

Somas
Cell bodies of neurons, containing the nucleus. They are considered a key processing part of the neuron.

Spatial light modulators
Components, for example, based on liquid crystals, used in optical computing to induce a spatially varying modulation on a beam of light.

Spikes
Short peaks of electrical potential at the membrane of a neuron, used to encode and communicate information. Also known as action potentials.

Spintronic
Spintronics is the field of study of systems in which information is encoded using the magnetic properties of electrons. The name is a contraction of 'spin' and 'electronics'.

possibly commercial — neuromorphic computing in the near future. Reaching the commercial level will require specific steps, on top of experimental and theoretical breakthroughs: optimization of technology for neuromorphic applications[147]; adaptation of computer-aided design tools to design neuromorphic hardware efficiently[148]; focusing on key applications in which neuromorphic computing has a clear advantage over conventional AI; and developing appropriate datasets, benchmarks and standards.

In parallel to the development of these new physical neural networks, solutions based on conventional digital CMOS are improving at a rapid pace, owing to the demand for fast application-oriented AI processors. More recent graphics processing units are co-designed to minimize data motion and reduce bit precision, thus achieving extremely low power consumption for specific neural network types and applications[149]. Evolving even faster is the field of AI, in which new algorithms and learning techniques are proposed at an extremely fast pace and immediately adopted by the community. A challenge and impetus for the physics-based neuromorphic computing community is to keep up with the advances on the two fronts of conventional CMOS chips and AI algorithms.

One approach that will help the physics-based community to keep pace is to build radically interdisciplinary teams that develop hand in hand the algorithms and the hardware. Another is to identify the challenges that current CMOS and AI algorithms will not be able to answer in the long term, and offer solutions to these issues based on new physics and materials. Such issues include technical challenges such as scaling-up and increasing the interconnectivity of neural networks, and conceptual challenges such as the possibility of performing lifelong learning, or being able to learn from few examples, as the brain does. These approaches are what we have highlighted in this Perspective article, especially stressing the importance to take inspiration from the blooming field of neuroscience, as well as AI.

Danijela Marković[1], *Alice Mizrahi*[1], *Damien Querlioz*[2]✉ *and Julie Grollier*[1]✉

[1]*Unité Mixte de Physique, CNRS, Thales, Université Paris-Saclay, Palaiseau, France.*

[2]*Université Paris-Saclay, CNRS, Centre de Nanosciences et de Nanotechnologies, Palaiseau, France.*

✉*e-mail: damien.querlioz@c2n.upsaclay.fr; julie.grollier@cnrs-thales.fr*

https://doi.org/10.1038/s42254-020-0208-2

Published online 28 July 2020

1. Lockery, S. R. The computational worm: spatial orientation and its neuronal basis in *C. elegans*. *Curr. Opin. Neurobiol.* **21**, 782–790 (2011).
2. French, R. M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **3**, 128–135 (1999).
3. Zenke, F., Poole, B. & Ganguli, S. Continual learning through synaptic intelligence. *Int. Conf. Mach. Learn.* **70**, 3987–3995 (2017).
4. Kirkpatrick, J. et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl Acad. Sci. USA* **114**, 3521–3526 (2017).
5. Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-inspired artificial intelligence. *Neuron* **95**, 245–258 (2017).
6. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
7. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
8. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
9. Chialvo, D. R. Emergent complex neural dynamics. *Nat. Phys.* **6**, 744–750 (2010).
10. Rabinovich, M. I., Varona, P., Selverston, A. I. & Abarbanel, H. D. I. Dynamical principles in neuroscience. *Rev. Mod. Phys.* **78**, 1213–1265 (2006).
11. Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge Univ. Press, 2014).
12. Sompolinsky, H., Crisanti, A. & Sommers, H. J. Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259–262 (1988).
13. Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top–down processing. *Nat. Rev. Neurosci.* **2**, 704–716 (2001).
14. Buzsaki, G. *Rhythms of the Brain* (Oxford Univ. Press, 2011).
15. McDonnell, M. D. & Ward, L. M. The benefits of noise in neural systems: bridging theory and experiment. *Nat. Rev. Neurosci.* **12**, 415–426 (2011).
16. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
17. Hoppensteadt, F. C. & Izhikevich, E. M. Oscillatory neurocomputers with dynamic connectivity. *Phys. Rev. Lett.* **82**, 2983–2986 (1999).
18. Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
19. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925–933 (2013).
20. Schliebs, S. & Kasabov, N. Evolving spiking neural network — a survey. *Evol. Syst.* **4**, 87–98 (2013).
21. Beyeler, M., Dutt, N. D. & Krichmar, J. L. Categorization and decision-making in a neurobiologically plausible spiking network using a STDP-like learning rule. *Neural Netw.* **48**, 109–124 (2013).
22. Dayan, P. & Abbott, L. F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2005).
23. Antle, M. C. & Silver, R. Orchestrating time: arrangements of the brain circadian clock. *Trends Neurosci.* **28**, 145–151 (2005).
24. Big data needs a hardware revolution. *Nature* **554**, 145–146 (2018).
25. Indiveri, G. et al. Neuromorphic silicon neuron circuits. *Front. Neurosci.* **5**, 73 (2011).
26. Furber, S. Large-scale neuromorphic computing systems. *J. Neural Eng.* **13**, 051001 (2016).
27. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
28. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
29. Xu, X. et al. Scaling for edge inference of deep neural networks. *Nat. Electron.* **1**, 216–222 (2018).
30. Ielmini, D. & Waser, R. (eds) *Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications* (Wiley, 2016).
31. Wouters, D. J., Waser, R. & Wuttig, M. Phase-change and redox-based resistive switching memories. *Proc. IEEE* **103**, 1274–1288 (2015).
32. Ha, S. D., Shi, J., Meroz, Y., Mahadevan, L. & Ramanathan, S. Neuromimetic circuits with synaptic devices based on strongly correlated electron systems. *Phys. Rev. Appl.* **2**, 064003 (2014).
33. Chanthbouala, A. et al. A ferroelectric memristor. *Nat. Mater.* **11**, 860–864 (2012).
34. Strukov, D. B. & Likharev, K. K. A reconfigurable architecture for hybrid CMOS/nanodevice circuits. *Proc. ACM/SIGDA Int. Symp. Field Progr. Gate Arrays* https://doi.org/10.1145/1117201.1117221 (2006).
35. Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
36. Bayat, F. M. et al. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits. *Nat. Commun.* **9**, 2331 (2018).
37. Narayanan, P. et al. Toward on-chip acceleration of the backpropagation algorithm using nonvolatile memory. *IBM J. Res. Dev.* **61**, 11:1–11:11 (2017).
38. Li, C. et al. Three-dimensional crossbar arrays of self-rectifying Si/SiO_2/Si memristors. *Nat. Commun.* 15666 (2017).
39. Ambrogio, S. et al. Neuromorphic learning and recognition with one-transistor-one-resistor synapses and bistable metal oxide RRAM. *IEEE Trans. Electron. Devices* **63**, 1508–1515 (2016).
40. Yao, P. et al. Face classification using electronic synapses. *Nat. Commun.* **8**, 15199 (2017).
41. Hirtzlin, T. et al. Digital biologically plausible implementation of binarized neural networks with differential hafnium oxide resistive memory arrays. *Front. Neurosci.* **13**, 1383 (2020).
42. Ji, Y. et al. Flexible and twistable non-volatile memory cell array with all-organic one diode–one resistor architecture. *Nat. Commun.* **4**, 2707 (2013).
43. Fuller, E. J. et al. Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. *Science* **364**, 570–574 (2019).
44. Noé, P. et al. Toward ultimate nonvolatile resistive memories: the mechanism behind ovonic threshold switching revealed. *Sci. Adv.* **6**, eaay2830 (2020).
45. Ambrogio, S. et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67 (2018).
46. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
47. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
48. Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
49. Larger, L. et al. Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing. *Opt. Express* **20**, 3241–3249 (2012).
50. Vandoorne, K. et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **5**, 3541 (2014).
51. Moughames, J. et al. Three dimensional waveguide-interconnects for scalable integration of photonic neural networks. *Optica* **7**, 640–646 (2020).
52. Heuser, T., Große, J., Kaganskiy, A., Brunner, D. & Reitzenstein, S. Fabrication of dense diameter-tuned quantum dot micropillar arrays for applications in photonic information processing. *APL Photon.* **3**, 116103 (2018).
53. Hayenga, W. E., Garcia-Gracia, H., Hodaei, H., Fainman, Y. & Khajavikhan, M. Metallic coaxial nanolasers. *Adv. Phys. X* **1**, 262–275 (2016).
54. Miscuglio, M. et al. All-optical nonlinear activation function for photonic neural networks [Invited]. *Opt. Mater. Express* **8**, 3851–3863 (2018).
55. Diehl, P. U., Zarrella, G., Cassidy, A., Pedroni, B. U. & Neftci, E. Conversion of artificial recurrent neural networks to spiking neural networks for low-power neuromorphic hardware. *Proc. IEEE Int. Conf. Reboot. Comput.* https://doi.org/10.1109/ICRC.2016.7738691 (2016).
56. Diehl, P. U. et al. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. *Proc. Int. Joint Conf. Neural Netw.* https://doi.org/10.1109/IJCNN.2015.7280696 (2015).
57. Esser, S. K., Appuswamy, R., Merolla, P., Arthur, J. V. & Modha, D. S. Backpropagation for energy-efficient neuromorphic computing. *Advances Neural Inform. Process. Systems* **28**, 1117–1125 (2015).

58. Poirazi, P., Brannon, T. & Mel, B. W. Pyramidal neuron as two-layer neural network. *Neuron* **37**, 989–999 (2003).
59. David, B., Idan, S. & Michael, L. Single cortical neurons as deep artificial neural networks. Preprint at *bioRxiv* https://doi.org/10.1101/613141 (2019).
60. Conrad, M., Engl, E. & Jolivet, R. B. Energy use constrains brain information processing. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM.2017.8268370 (2017).
61. Gidon, A. et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* **367**, 83–87 (2020).
62. London, M. & Häusser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
63. Lenk, K. et al. A computational model of interactions between neuronal and astrocytic networks: the role of astrocytes in the stability of the neuronal firing rate. *Front. Comput. Neurosci.* **13**, 92 (2020).
64. Mead, C. & Ismail, M. (eds) *Analog VLSI Implementation of Neural Systems* (Springer, 1989).
65. Boahen, K. A neuromorph's prospectus. *Comput. Sci. Eng.* **19**, 14–28 (2017).
66. Arthur, J. V. & Boahen, K. A. Silicon-neuron design: a dynamical systems approach. *IEEE Trans. Circuits Syst. I Regul. Pap.* **58**, 1034–1043 (2011).
67. Ohno, T. et al. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591–595 (2011).
68. La Barbera, S., Vuillaume, D. & Alibart, F. Filamentary switching: synaptic plasticity through device volatility. *ACS Nano* **9**, 941–949 (2015).
69. Stoliar, P. et al. A leaky-integrate-and-fire neuron analog realized with a mott insulator. *Adv. Funct. Mater.* **27**, 1604740 (2017).
70. Valle, Jdel et al. Subthreshold firing in Mott nanodevices. *Nature* **569**, 388–392 (2019).
71. Pickett, M. D., Medeiros-Ribeiro, G. & Williams, R. S. A scalable neuristor built with Mott memristors. *Nat. Mater.* **12**, 114–117 (2013).
72. Kumar, S., Strachan, J. P. & Williams, R. S. Chaotic dynamics in nanoscale NbO_2 Mott memristors for analogue computing. *Nature* **548**, 318–321 (2017).
73. Parihar, A., Shukla, N., Jerry, M., Datta, S. & Raychowdhury, A. Computational paradigms using oscillatory networks based on state-transition devices. *Proc. Int. Joint Conf. Neural Netw.* https://doi.org/10.1109/IJCNN.2017.7966285 (2017).
74. Sharma, A. A., Bain, J. A. & Weldon, J. A. Phase coupling and control of oxide-based oscillators for neuromorphic computing. *IEEE J. Explor. Solid State Comput. Devices Circuits* **1**, 58–66 (2015).
75. Li, S., Liu, X., Nandi, S. K., Venkatachalam, D. K. & Elliman, R. G. High-endurance megahertz electrical self-oscillation in Ti/NbO_x bilayer structures. *Appl. Phys. Lett.* **106**, 212902 (2015).
76. Yi, W. et al. Biological plausibility and stochasticity in scalable VO_2 active memristor neurons. *Nat. Commun.* **9**, 4661 (2018).
77. Fell, J. & Axmacher, N. The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* **12**, 105–118 (2011).
78. Ignatov, M., Ziegler, M., Hansen, M. & Kohlstedt, H. Memristive stochastic plasticity enables mimicking of neural synchrony: memristive circuit emulates an optical illusion. *Sci. Adv.* **3**, e1700849 (2017).
79. Arnaud, F. et al. Truly Innovative 28nm FDSOI technology for automotive micro-controller applications embedding 16MB phase change memory. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM.2018.8614595 (2018).
80. Suri, M. et al. Phase change memory as synapse for ultra-dense neuromorphic systems: application to complex visual pattern extraction. *Proc. Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM.2011.6131488 (2011).
81. Boybat, I. et al. Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **9**, 2514 (2018).
82. Tuma, T., Pantazi, A., Gallo, M. L., Sebastian, A. & Eleftheriou, E. Stochastic phase-change neurons. *Nat. Nanotechnol.* **11**, 693–699 (2016).
83. Boyn, S. et al. Learning through ferroelectric domain dynamics in solid-state synapses. *Nat. Commun.* **8**, 14736 (2017).
84. Oh, S., Hwang, H. & Yoo, I. K. Ferroelectric materials for neuromorphic computing. *APL Mater.* **7**, 091109 (2019).
85. Alzate, J. G. et al. 2 MB array-level demonstration of STT-MRAM process and performance towards L4 cache applications. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM19573.2019.8993474 (2019).
86. Vansteenkiste, A. et al. The design and verification of MuMax3. *AIP Adv.* **4**, 107133 (2014).
87. Grollier, J. et al. Neuromorphic spintronics. *Nat. Electron.* https://doi.org/10.1038/s41928-019-0360-9 (2020).
88. Torrejon, J. et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428–431 (2017).
89. Borders, W. A. et al. Integer factorization using stochastic magnetic tunnel junctions. *Nature* **573**, 390–393 (2019).
90. Burgt, Yvande et al. A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mater.* **16**, 414–418 (2017).
91. Pecqueur, S. et al. Neuromorphic time-dependent pattern classification with organic electrochemical transistor arrays. *Adv. Electron. Mater.* **4**, 1800166 (2018).
92. Fon, W. et al. Complex dynamical networks constructed with fully controllable nonlinear nanomechanical oscillators. *Nano Lett.* **17**, 5977–5983 (2017).
93. Coulombe, J. C., York, M. C. A. & Sylvestre, J. Computing with networks of nonlinear mechanical oscillators. *PLoS ONE* **12**, e0178663 (2017).
94. Likharev, K. K. & Semenov, V. K. RSFQ logic/memory family: a new Josephson-junction technology for sub-terahertz-clock-frequency digital systems. *IEEE Trans. Appl. Supercond.* **1**, 3–28 (1991).
95. Russek, S. E. et al. Stochastic single flux quantum neuromorphic computing using magnetically tunable Josephson junctions. *Proc. IEEE Int. Conf. Reboot. Comput.* https://doi.org/10.1109/ICRC.2016.7738712 (2016).
96. Schneider, M. L. et al. Ultralow power artificial synapses using nanotextured magnetic Josephson junctions. *Sci. Adv.* **4**, e1701329 (2018).
97. Wang, M. et al. Robust memristors based on layered two-dimensional materials. *Nat. Electron.* **1**, 130–136 (2018).
98. Shi, Y. et al. Electronic synapses made of layered two-dimensional materials. *Nat. Electron.* **1**, 458–465 (2018).
99. Chaudhuri, R. & Fiete, I. Computational principles of memory. *Nat. Neurosci.* **19**, 394–403 (2016).
100. Romeira, B., Avó, R., Figueiredo, J. M. L., Barland, S. & Javaloyes, J. Regenerative memory in time-delayed neuromorphic photonic resonators. *Sci. Rep.* **6**, 1–12 (2016).
101. Appeltant, L. et al. Information processing using a single dynamical node as complex system. *Nat. Commun.* **2**, 468 (2011).
102. Larger, L. et al. High-speed photonic reservoir computing using a time-delay-based architecture: million words per second classification. *Phys. Rev. X* **7**, 011015 (2017).
103. Antonik, P., Haelterman, M. & Massar, S. Brain-inspired photonic signal processor for generating periodic patterns and emulating chaotic systems. *Phys. Rev. Appl.* **7**, 054014 (2017).
104. Antonik, P., Marsal, N., Brunner, D. & Rontani, D. Human action recognition with a large-scale brain-inspired photonic computer. *Nat. Mach. Intell.* **1**, 530–537 (2019).
105. Soudry, D., Castro, D. D., Gal, A., Kolodny, A. & Kvatinsky, S. Memristor-based multilayer neural networks with online gradient descent training. *IEEE Trans. Neural Netw. Learn. Syst.* **26**, 2408–2421 (2015).
106. Yu, S. Neuro-inspired computing with emerging nonvolatile memorys. *Proc. IEEE* **106**, 260–285 (2018).
107. Lastras-Montaño, M. A. & Cheng, K.-T. Resistive random-access memory based on ratioed memristors. *Nat. Electron.* **1**, 466–472 (2018).
108. Shi, Y. et al. Adaptive quantization as a device-algorithm co-design approach to improve the performance of in-memory unsupervised learning with SNNs. *IEEE Trans. Electron. Devices* **66**, 1722–1728 (2019).
109. Hirtzlin, T. et al. Outstanding bit error tolerance of resistive ram-based binarized neural networks. *Proc. IEEE Int. Conf. Artificial Intell. Circuits Systems* https://doi.org/10.1109/AICAS.2019.8771544 (2019).
110. Lin, X., Zhao, C. & Pan, W. Towards accurate binary convolutional neural network. *Advances Neural Inform. Process. Systems* **30**, 345–353 (2017).
111. Penkovsky, B. et al. In-memory resistive ram implementation of binarized neural networks for medical applications. *Proc. IEEE Process. Design Automat. Test Europe Conf.* https://doi.org/10.23919/DATE48585.2020.9116439 (2020).
112. Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R. & Bengio, Y. Binarized neural networks. *Advances Neural Inform. Process. Systems* **29**, 4107–4115 (2016).
113. Rastegari, M., Ordonez, V., Redmon, J. & Farhadi, A. XNOR-net: ImageNet classification using binary convolutional neural networks. *Comput. Vision* **4**, 525–542 (2016).
114. Hirtzlin, T. et al. Hybrid analog-digital learning with differential RRAM synapses. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM19573.2019.8993555 (2019).
115. Shi, Y. et al. Neuroinspired unsupervised learning and pruning with subquantum CBRAM arrays. *Nat. Commun.* **9**, 5312 (2018).
116. Shi, Y., Nguyen, L., Oh, S., Liu, X. & Kuzum, D. A soft-pruning method applied during training of spiking neural networks for in-memory computing applications. *Front. Neurosci.* **13**, 405 (2019).
117. Ernoult, M., Grollier, J. & Querlioz, D. Using memristors for robust local learning of hardware restricted Boltzmann machines. *Sci. Rep.* **9**, 1851 (2019).
118. Ishii, M. et al. On-chip trainable 1.4M 6T2R PCM synaptic array with 1.6K stochastic LIF neurons for spiking RBM. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM19573.2019.8993466 (2019).
119. Querlioz, D., Bichler, O., Dollfus, P. & Gamrat, C. Immunity to device variations in a spiking neural network with memristive nanodevices. *IEEE Trans. Nanotechnol.* **12**, 288–295 (2013).
120. Bill, J. & Legenstein, R. A compound memristive synapse model for statistical learning through STDP in spiking neural networks. *Neuromorphic Eng.* **8**, 412 (2014).
121. Querlioz, D., Bichler, O., Vincent, A. F. & Gamrat, C. Bioinspired programming of memory devices for implementing an inference engine. *Proc. IEEE* **103**, 1398–1416 (2015).
122. Bi, G.-Q. & Poo, M.-M. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* **24**, 139–166 (2001).
123. Jo, S. H. et al. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
124. Kim, S. et al. Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett.* **15**, 2203–2211 (2015).
125. Barbera, S. L., Vincent, A. F., Vuillaume, D., Querlioz, D. & Alibart, F. Interplay of multiple synaptic plasticity features in filamentary memristive devices for neuromorphic computing. *Sci. Rep.* **6**, 39216 (2016).
126. Serb, A. et al. Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. *Nat. Commun.* **7**, 12611 (2016).
127. Pedretti, G. et al. Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity. *Sci. Rep.* **7**, 5288 (2017).
128. Srinivasan, G. & Roy, K. ReStoCNet: residual stochastic binary convolutional spiking neural network for memory-efficient neuromorphic computing. *Front. Neurosci.* **13**, 189 (2019).
129. Mozafari, M., Kheradpisheh, S. R., Masquelier, T., Nowzari-Dalini, A. & Ganjtabesh, M. First-spike-based visual categorization using reward-modulated STDP. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 6178–6190 (2018).
130. Mizrahi, A. et al. Controlling the phase locking of stochastic magnetic bits for ultra-low power computation. *Sci. Rep.* **6**, 30535 (2016).
131. Dalgaty, T., Castellani, N., Querlioz, D. & Vianello, E. In-situ learning harnessing intrinsic resistive memory variability through Markov chain Monte Carlo sampling. Preprint at https://arxiv.org/abs/2001.11426 (2020).
132. Pinna, D. et al. Skyrmion gas manipulation for probabilistic computing. *Phys. Rev. Appl.* **9**, 064018 (2018).
133. Mizrahi, A. et al. Neural-like computing with populations of superparamagnetic basis functions. *Nat. Commun.* **9**, 1533 (2018).
134. Romera, M. et al. Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).
135. Wang, Z. et al. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **1**, 137–145 (2018).

136. Türel, Ö., Lee, J. H., Ma, X. & Likharev, K. K. Neuromorphic architectures for nanoelectronic circuits. *Int. J. Circ. Theor. Appl.* **32**, 277–302 (2004).
137. Demis, E. C. et al. Atomic switch networks — nanoarchitectonic design of a complex system for natural computing. *Nanotechnology* **26**, 204003 (2015).
138. Neckar, A. et al. Braindrop: a mixed-signal neuromorphic architecture with a dynamical systems-based programming model. *Proc. IEEE* **107**, 144–164 (2019).
139. Fujii, K. & Nakajima, K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys. Rev. Appl.* **8**, 024030 (2017).
140. Yamamoto, Y. et al. Coherent Ising machines — optical neural networks operating at the quantum limit. *npj Quantum Inf.* **3**, 1–15 (2017).
141. Tacchino, F., Macchiavello, C., Gerace, D. & Bajoni, D. An artificial neuron implemented on an actual quantum processor. *npj Quantum Inf.* **5**, 1–8 (2019).
142. Mochida, R. et al. A 4M synapses integrated analog ReRAM based 66.5 TOPS/W neural-network processor with cell current controlled writing and flexible network architecture. *Proc. IEEE Symp. VLSI Technology* https://doi.org/10.1109/VLSIT.2018.8510676 (2018).
143. Ishii, M. et al. On-chip trainable 1.4M 6T2R PCM synaptic array with 1.6K stochastic LIF neurons for spiking RBM. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM19573.2019.8993466 (2019).
144. Liu, Q. et al. A fully integrated analog ReRAM based 78.4TOPS/W compute-in-memory chip with fully parallel MAC computing. *Proc. IEEE Int. Solid-State Circuits Conf.* https://doi.org/10.1109/ISSCC19947.2020.9062953 (2020).
145. Golonzka, O. et al. Non-volatile RRAM embedded into 22FFL FinFET technology. *Proc. Symp. VLSI Technology* https://doi.org/10.23919/VLSIT.2019.8776570 (2019).
146. Golonzka, O. et al. MRAM as embedded non-volatile memory solution for 22FFL FinFET technology. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM.2018.8614620 (2018).
147. Ambrogio, S. et al. Reducing the impact of phase-change memory conductance drift on the inference of large-scale hardware neural networks. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM19573.2019.8993482 (2019).
148. Chen, P.-Y., Peng, X. & Yu, S. NeuroSim+: an integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures. *Proc. IEEE Int. Electron Devices Meeting* https://doi.org/10.1109/IEDM.2017.8268337 (2017).
149. Dally, W. J. et al. Hardware-enabled artificial intelligence. *Proc. IEEE Symp. VLSI Circuits* https://doi.org/10.1109/VLSIC.2018.8502368 (2018).
150. Caulfield, H. J. & Dolev, S. Why future supercomputing requires optics. *Nat. Photon.* **4**, 261–263 (2010).
151. Tucker, R. S. The role of optics in computing. *Nat. Photon.* **4**, 405 (2010).
152. Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133–1145 (2001).
153. Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for modern deep learning research. *AAAI* **34**, 13693–13696 (2019).
154. Nvidia AI. BERT meets GPUs. *Medium* https://medium.com/future-vision/bert-meets-gpus-403d3fbed848 (2020).
155. Schneidman, E., Freedman, B. & Segev, I. Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural Comput.* **10**, 1679–1703 (1998).
156. Branco, T., Staras, K., Darcy, K. J. & Goda, Y. Local dendritic activity sets release probability at hippocampal synapses. *Neuron* **59**, 475–485 (2008).
157. Harris, J. J., Jolivet, R., Engl, E. & Attwell, D. Energy-efficient information transfer by visual pathway synapses. *Curr. Biol.* **25**, 3151–3160 (2015).

Acknowledgements
This work was supported as part of the Q-MEEN-C, an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under award DE-SC0019273 and by the European Research Council ERC under grant bioSPINspired (682955) and NANOINFER (715872). A.M. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 824103 (NEUROTECH).

Author contributions
All authors wrote the Perspective article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Physics thanks Wei Lu and the other, anonymous, reviewers for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.