Adaptive Extreme Edge Computing for Wearable Devices

Adaptive Extreme Edge Computing for Wearable Devices

Erika Covi[1,*], **Elisa Donati**[2,*], **Hadi Heidari**[3,*], **David Kappel**[4,*], **Xiangpeng Liang**[3,*], **Melika Payvand**[2,*], **and Wei Wang**[5,*]

[1]NaMLab gGmbH, Nöthnitzer Strasse 64 a, 01187 Dresden, Germany
[2]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland
[3]Microelectronics Lab (meLAB), James Watt School of Engineering, University of Glasgow, G12 8QQ, UK
[4]Bernstein Center for Computational Neuroscience, III Physikalisches Institut - Biophysik, Georg-August Universität, Göttingen, Germany
[5]The Andrew and Erna Viterbi Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel,Formerly with Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano and IU.NET, Milan, Italy
[*]All authors contributed equally to this work

ABSTRACT

Wearable devices are a fast-growing technology with impact on personal healthcare for both society and economy. Due to the widespread of sensors in pervasive and distributed networks, power consumption, processing speed, and system adaptation are vital in future smart wearable devices. The visioning and forecasting of how to bring computation to the edge in smart sensors have already begun, with an aspiration to provide adaptive extreme edge computing. Here, we provide a holistic view of hardware and theoretical solutions towards smart wearable devices that can provide guidance to research in this pervasive computing era. We propose various solutions for biologically plausible models for continual learning in neuromorphic computing technologies for wearable sensors. To envision this concept, we provide a systematic outline in which prospective low power and low latency scenarios of wearable sensors in neuromorphic platforms are expected. We successively describe vital potential landscapes of neuromorphic processors exploiting complementary metal-oxide semiconductors (CMOS) and emerging memory technologies (e.g. memristive devices). Furthermore, we evaluate the requirements for edge computing within wearable devices in terms of footprint, power consumption, latency, and data size. We additionally investigate the challenges beyond neuromorphic computing hardware, algorithms and devices that could impede enhancement of adaptive edge computing in smart wearable devices.

Keywords: Neuromorphic computing, Edge computing, Wearable devices, Learning algorithms, Memristive devices

1 Introduction

Wearable devices can monitor various human body symptoms ranging from heart, respiration, movement, to brain activities. Such miniaturized devices using different sensors can detect, predict, and analyze the physical performance, physiological status, biochemical composition, and mental alertness of the human body. Despite advances in novel materials that can improve the resolution and sensitivity of sensors, modern wearable devices are facing various challenges such as low computing capability, high power consumption, high amount of data to be transmitted, and low speed of the data transmission. Conventional wearable sensing solutions mostly transmit the collected data to external servers for off-chip computing and processing. This approach typically creates an information bottleneck acting as one of the major limiting factors in lowering the power consumption and improving the speed of the operation of the sensing systems. In addition, the use of conventional remote servers with conventional signal processing techniques for processing these temporal real-time sensing data makes it computationally intensive and results in significant power consumption and hardware occupation. Moreover, standard von-Neumann architectures feature a physical separation between memory and processing unit, thus further increasing the power consumption to shuttle data between units. Such solutions always need a trade-off between power lifetime and computing capability. Bringing computing at the edge enables faster response times and opens the possibility of personalized always-on wearable devices able for continuously interacting and learning with the environment. However, a radical change of paradigm which uses innovative algorithms, circuits and memory devices is needed to maximize the system performance whilst keeping power and memory budgets at a minimum.

Conventional computers, using Boolean and bit-precise digital representations and executing operations with time-

multiplexed and clocked signal, are not optimized for fuzzy inputs and complex cognitive tasks such as pattern recognition, time series prediction, and decision making. Deep Artificial Neural Networks (ANNs) on the other hand have demonstrated amazing results in a wide range of pattern recognition tasks including machine vision, Natural Language Processing (NLP), and speech recognition[1,2]. Dedicated hardware ANN accelerators, including Graphical Processing Units (GPUs), Tensor Processing Units (TPUs), and custom Application Specific Integrated Circuits (ASICs) with parallel architectures are being developed to execute these algorithms and obtain high accuracy inference results. GPUs provide a substrate for parallel processing nature of the ANNs and thanks to its very long memory bus it is perfect for running Vector Matrix Multiplications (VMMs) which are at the core of the processing in deep neural networks. Therefore, GPUs support the parallelism whose massive version exists in the brains for cognitive purposes, but they consume orders of magnitude more power than that of the brain[3], since they are clocked and the memory access is not localized. To solve this problem, ASIC accelerators try to reduce the complexity of the structure by making the system more application specific and using clock gating and specific hardware structure which matches best to the structure of the mapped neural network to reduce power consumption through less memory read and data access[4-7].

To go even further in power savings, there are two problems to be solved: (i) remove clock and (ii) perform computation with co-localization of memory and processor. The first problem calls for the development of event-based systems, where processing is performed "asynchronously", i.e. only when there are input "events". The algorithmic basis for this kind of "asynchronous" processing is Spiking Neural Network (SNN), in which neurons spike asynchronously only to communicate information to each other.

To avoid the data movement between the memory and the processor, the memory element should be not only used to store data but also to perform computation inside the processor. This approach is called "in-memory computing". These two approaches of (i) event-based systems and (ii) in-memory computing, together with (iii) massive parallelism, are the three fundamental principles which have led to the development of neuromorphic computing, and to the realization of highly efficient neuromorphic platforms[8-13]. Therefore, in this article, we will refer to event-based highly parallel systems that are able to perform real-time sensory processing.

Despite that current fully Complementary Metal-Oxide-Semiconductor (CMOS) implementations of neuromorphic platforms have shown remarkable performance in terms of power efficiency and classification accuracy, there are still some bottlenecks hindering the design of embedded sensing and processing systems. First, the memory used is typically Static Random Access Memory (SRAM), which has very low static power consumption, but it is a large element (6 transistors per cell) and it is volatile. The latter feature implies that the information about the network configuration has to be stored elsewhere and transferred to the system at its startup. For large networks, it may take tens of minutes before the system is ready for normal operation. Second, always-on adaptive systems need to work with time constants that have the same time-span of the task that is being learned (e.g. longer than seconds). Implementing such long time constants in neuromorphic CMOS circuits is impractical, since it requires large area capacitors.

To overcome the limitations of fully CMOS-based approaches, the intrinsic unique physical properties of emerging memristive devices can be exploited for both long-term (non-volatile) weight storage and short-term (volatile) task-relevant timescales. In particular, non-volatile devices feature retention times on a long time scale (>10 years,[14-17]) while showing weight reconfigurability with voltages compatible with typical CMOS circuits ($\leq$ 3.3 V). Volatile devices, instead, can have time constants in the order of tens of milliseconds to seconds ([18-23]), thus being able to emulate biological time constants. This non-volatile / volatile property of memristive devices, together with a small footprint and power efficiency, has indeed attracted a lot of interest in the last ten years[24-26]. However, memristive technology has to be supported by *ad hoc* theoretically sound biologically plausible algorithms enabling continual learning and capable to exploit the intrinsic physical properties of memristive devices, such as stochasticity, to achieve accuracy performance comparable to state-of-the-art ANN whilst reducing the power consumption.

This review discusses the challenges to undertake for designing extreme edge computing wearable devices in four different categories: (i) the state-of-the-art wearable sensors and main restrictions towards low-power and high performance learning capabilities; (ii) different algorithms for modeling biologically plausible continual learning; (iii) CMOS-based neuromorphic processors and signal processing techniques enabling low-power local edge computing strategies; (iv) emerging memristive devices for more efficient and scalable embedded intelligent systems. As graphically summarized in Fig. 1, we argue that a holistic approach which combines and exploits all the strengths of these four categories in a co-designed system is the key factor enabling future generations of smart sensing systems.

2 Wearable sensors

Sensors act as the information collector of a machine or a system that can respond to its physical ambient environment. They are able to translate a specific type of information from a physical environment such as the human body to an electrical signal ([27]). For collecting the information from the human body environment, wearable versions of the machine or the system, i.e. wearable devices, would be of great convenient and helpful. Wearable devices require miniaturize, flexible, and highly sensitive
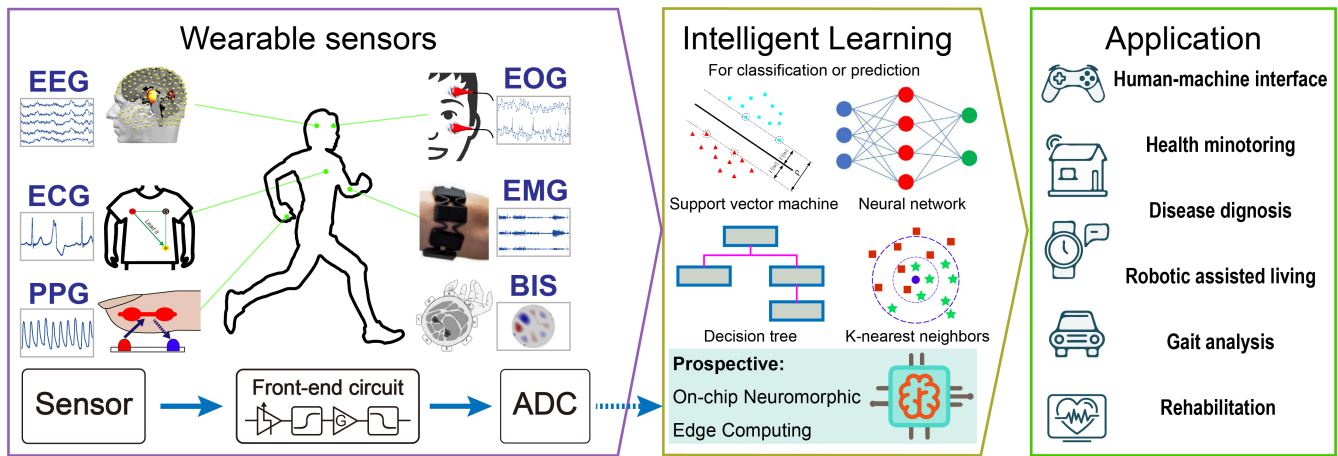
Figure 1. A graphical overview of adaptive edge computing in wearable biomedical devices. The figure shows the pathway from wearable sensors to their application through intelligent learning.

sensors to capture clear information from the body. However, from processing aspect and to make a signal meaningful towards personalized devices, further development is still needed.

Due to the fact that the sensing signal is relatively weak and noisy, a readout circuit (normally composed by an amplifier, a conditioning circuit and an analogue signal processing unit) is necessary to make the signal readable for a system ([27,28]). The subsequent high-level system will process the data and send commands to actuators for a closed-loop control or interaction ([29–31]). For various applications ranging from the human-machine interface ([29]) to health monitoring ([32,33]), different combinations of sensor and system have been developed over the past decade ([34,35]). The use of machine learning empowers sensor to build a novel smart application. The examples will be provided in the next section.

2.1 Wearable sensors with machine learning

Recently, the field of artificial intelligence further boosts the possibility of smart wearable sensory systems. The emerging intelligent applications and high-performance systems require more complexity and demand sensory units accurately describe the physical object. The decision-making unit or algorithm can therefore output a more reliable result ([35–39]). Depending on the signal acquiring position, Fig. 1 summaries the four biopotential sensors and two widely used wearable sensors along with their learning systems and applications. The sensors for the biopotential will be introduced first, and the other two wearable sensors will be provided separately.

The biopotential signal can be extracted from the human body using a sensor with direct electrode contact. The electrochemical activity of the cells in nervous, muscular and glandular tissue generates ionic currents in the body. An electrode-electrolyte transducer is needed to convert the ionic current to electric current for the front-end circuit. The electrode that is normally made up of mental can be oxidized by the electrolyte, generating metal ions and free electrons. In addition, the anions in the electrolyte can also be oxidized to neutral atoms and free electrons. These free electrons result in current flow through the electrode. Thus, the surface potential generated by the electrochemical activities in cells can be sensed by the electrode. However, the bio-signals sensed by the electrode are weak and noisy. Before digitizing the collected signals by analog-to-digital converter, an analogue front-end is essential to provide a readable signal. The design requirements of the front-end for the biopotential electrodes can be summarized as follow: i) high common mode rejection ratio; ii) high signal-to-noise-ratio; iii) low-power consumption; iv) signal filtering, and v) configurable gain ([40]).

Electrocardiography (ECG). ECG is the electrical activity generated by the electrochemistry around cardiac tissue. Containing morphological or statistical features, ECG provides comprehensive information for analyzing and diagnosing cardiovascular diseases ([41]). In the previous study, automatic ECG classification has been achieved using machine learning algorithms, such as Deep Neural Network (DNN) ([42,43]), Support Vector Machine (SVM) ([44,45]), and Recurrent Neural Network (RNN) ([46,47]). According to Association for the Advancement of Medical Instrumentation, there are five classes of ECG type of interest: normal, ventricular, supraventricular, fusion of normal and ventricular, and unknown beats. These methodologies can be evaluated by available ECG database and yield over 90% accuracy and sensitivity for the five classes, which is essential for future cardiovascular health monitoring. In wearable application,[48] and[49] present systems that measure ECG and send it to the cloud for classification and health monitoring.

Electroencephalography (EEG). Our brain neurons communicate with each other through electrical impulses. An EEG electrode can help to detect potential information associated with this activity through investigating EEG ([50,51]) in the surface of the skull. In comparison with other biopotential signals, surface EEG is relatively weak (normally in the range of microvolt-level) and noisy ([52,53]). Therefore, it requires high input impedance readout circuit and intensive signal pre-processing for clean EEG data ([40,50]). While wet-electrode (Ag/AgCl) is more precise and more suitable for clinical purpose, passive dry-electrode is more suitable for daily health monitoring and brain-machine interface ([52,54]). Besides, the applications also include mental disorder ([55]), driving safety ([51,54]), and emotion evaluation ([56]). A commercial biopotential data acquisition system, Biosemi Active Two, provides up to 256 channels for EEG analysis ([57]). For a specific application, we can reduce the number of electrodes to only detect the relevant areas, such as 19 channels for depression diagnosis ([58]), four channels for evaluating driver vigilance ([51]) and 64 channels for emotional state classification ([56]). Although EEG is on-body biopotential, most of the existing EEG researches employed offline learning and analysis because of the system complexity and the high number of channels. In wearable real-time applications, usually a smaller number of channels were selected and the data were wirelessly sent to cloud for further processing ([51,54,59,60]).

Electrooculography (EOG). The eye movement, which results in potential variations around eyes as EOG, is a combined effect of environmental and psychological changes. It returns relatively weak voltage (0.01-0.1mV) and low frequency (0-10Hz) ([53]). Differ from other eye tracking techniques using a video camera and infrared, EOG provides a lightweight, inexpensive and fully wearable solution to access human's eye movement ([61]). It is the most widely used approach of wearable human-machine interface, especially for assisting quadriplegics ([61]). It has been used to control a wheelchair ([62]), control a prosthesis limb ([63]),([31]) evaluate sleeping ([64–66]). Additionally, recent studies fuse EEG and EOG to increase the degree of freedom of signal and enhance the system reliability because their similar implicit information such as sleepiness ([64,67]) and mental health ([68]). EOG can also act as a supplement to provide additional functions or commands to an EEG system ([31,69,70]).

Electromyography (EMG). EMG is an electrodiagnostic method for recording and analyzing the electrical activity generated by skeletal muscles. EMG is generated by skeletal muscle movement, which frequently occurs in arms and legs. It yields higher amplitude (up to 10 millivolts) and bandwidth (20-1000Hz) compared to the other biopotentials ([40,53]). Near the active muscle, different oscillation signals can be measured by a dry electrode array, which allows the computer to sense and decode body motion ([71–73]). A prime example is the Myo armband of Thalmic Labs, which is a commercial multi-sensor device that consists of EMG sensors, gyroscope, accelerometer and magnetometer ([74]). The sensory data is sent to phone or PC via Bluetooth, at which various body movements can be obtained by feature extraction and machine learning. Moreover, the application of EMG is frequently linked to target control like a wheelchair ([75]) and prosthetic hand ([76,77]) for assisting disabled people. In addition, its application also includes sign language recognition ([71]), diagnosis of neuromuscular disorders ([72,78]), analysis of walking strides ([73]) and virtual reality ([79]). Machine learning enables the system to overcome the variation of EMG signals from different users ([71,72]).

Photoplethysmography (PPG). PPG is an non-invasive and low-cost optical measurement method that is often used for blood pressure and heart rate monitoring in wearable devices. The optical properties in skin and tissue are periodically changes due to the blood flow driven by the heartbeat. By using a light emitter toward the skin surface, the photosensor can detect the variations in light absorption normally from wrist or finger. This variation signal is called PPG which is highly relevant to the rhythm of the cardiovascular system ([80]). Compared with ECG, PPG is easily accessible and low cost, which makes it an ideal intermedia of wearable heart rate measurement. The main disadvantage against ECG is that the PPG is not unique for different persons and body positions. Thus, further analysis of PPG requires machine learning or other statistics tools for calibrating the signal to different scenarios. For example, it can be used in biometric identification after deep learning ([81,82]). It is worth mentioning that PPG is a strong supplementary in the application of ECG.

Bioimpedance spectroscopy (BIS). BIS is another low-cost and powerful sensing technique that provides informative body parameters. The principle is that cell membrane behaves like a frequency-dependent capacitor and impedance. The emitter electrodes generate multifrequency excitation signal (0.1-100MHz) on the skin while the receiver electrodes collect these current for demodulating the impedance spectral data of the tissue in between ([83,84]). Compared to homogeneous materials, body tissue presents more complicated impedance spectra because of the cell membranes and macromolecules. Therefore, the tissue conditions, such as muscle concentration, structural and chemical composition, can be analysed through BIS. The BIS can measure body composition such as fat and water ([84]). Based on the different setup in terms of position and frequency, it can also be helpful in the early detection of diseases such as lymphedema, organ ischemia and cancer ([85]). Furthermore, multiple pair-wise electrodes can form electrical impedance tomography that describes impedance distribution. By embedding these electrodes in a wristband, the tomography can estimate hand gesture after training, which is another novel solution of inexpensive human-machine interface ([86]).

2.2 Multisensory fusion in wearable devices

Every sensor has its own limitation. In some demanding cases, an individual sensor itself cannot satisfy the system requirement such as accuracy or robustness ([35, 87–89]). The solution involves increasing the number and type of sensors to form a multisensory system or sensor network for one measurement purpose([87–89]). Multiple types of sensor synergistically working in a system provide more dimensions of input to fully map an object onto the data stream. Different sensors return different data with respect to sampling rate, number of input and the information behind the data. Machine learning models, such as ANN and SVM, can be designed to combine multiple sources of data. Depended on the application, sensor types and data structure, several approaches have been proposed for multisensory fusion. Generally, in such a system, machine learning is frequently used and plays an vital role in merging different sources of sensory data based on its multidimensional data processing mechanism. The machine learning algorithms allow sensory fusion occurs at the signal, feature or decision level([88, 89]). The results showed that a multisensory system is advantageous in improving system performance. For example, the fusion of ECG and PPG pattern can be an informative physiological parameter for robust medical assessment ([90]). Counting the peak intervals between PPG and ECG can estimate the arterial blood pressure ([91]). Interestingly, a recent study shows that the QRS complex of ECG can be reconstructed from PPG by a novel transformed attentional neural networks after training ([92]). This could be beneficial for the accessibility of wearable ECG.

2.3 Challenges towards smart wearable sensors with edge computing

Given the potential of the sensory system with machine learning, the main challenge raised is the shortage of power and computing efficient ([28]). The novel applications using multiple sensors and high learning ability usually require more energy in the wearable computing unit ([33]). Nevertheless, the power supply in the wearable domain is a difficulty with existing battery technologies. This weakness limits the further development of smart wearable device ([33]). The existing solution is to wirelessly transfer the raw data onto a cloud where the computationally intensive algorithm is implemented ([93]). However, this solution is not ideal considering 1) the complexity of using a wireless module, 2) the non-negligible power consumption, 3) the amount of data, 4) the space limitation due to the range of wireless transmission, 5) privacy issues due to the broadcast of signals, 6) non-negligible time latency due to communication channel. These drawbacks strongly limit the application of wearable sensors.

Implementation of ANN in von Neumann architectures, which has been frequently used in sensors, is power-hungry. Conversely, it has been reported that signal processing activity in the brain is several orders of magnitudes more power-efficient and one order in processing rate better than digital systems ([94]). Compared to conventional approaches based on a binary digital system, brain-inspired neuromorphic hardware yet to be advanced in the contexts of data storage and removal as well as their transmission between different units. In this perspective, a neuromorphic chip with a built-in intelligent algorithm can act as a front-end processor next to the sensor. The conventional Analog to Digital Converters (ADCs) could be replaced by a delta encoder or feature extractor converting the sensor analog output to spike-based signal for the hardware (see Section 4). In the end, the output becomes the result of recognition or prediction instead of an intensive data stream. In this way, the computation occurs at the local edge under low power and brain-like architecture.

3 Models for biologically plausible continual learning

In this section we will highlight some recently introduced methods to port the power of modern machine learning to neuromorphic edge devices. In the last couple of years, machine learning has made big steps forward reaching close-to human performance on a wide range of tasks. Many of the most successful machine learning methods are based on artificial neural networks (ANN), which are inspired by the organization of information processing in the brain. However – somewhat contradictory – mapping modern ANN learning methods to brain-inspired hardware poses considerable challenges to the algorithm and hardware design. The main reason for this is, that the development of machine learning algorithms has been strongly influenced by the development of powerful mainframe computers that perform learning offline in big server farms only eventually sending back results to the user. While this development has paved the ground for today's success of ANNs, it has also lead the field away from following the principles used in biology for efficient learning. In the following Section 3.1 we will review recent approaches to combine the strengths of modern machine learning and brain-inspired algorithms, that are of particular interest for edge computing applications. In Section 3.2 we will focus on the problem to cope with extreme memory constraints by exploiting sparsity. In Section 3.3 we will highlight additional open challenges and future work.

3.1 Brain-inspired learning algorithms for neuromorphic hardware

Today, the dominating method for training artificial neural networks is the error backpropagation (Backprop) algorithm[100], which provides an efficient and scalable solution to adapting the network parameters to a set of training data. Backprop is
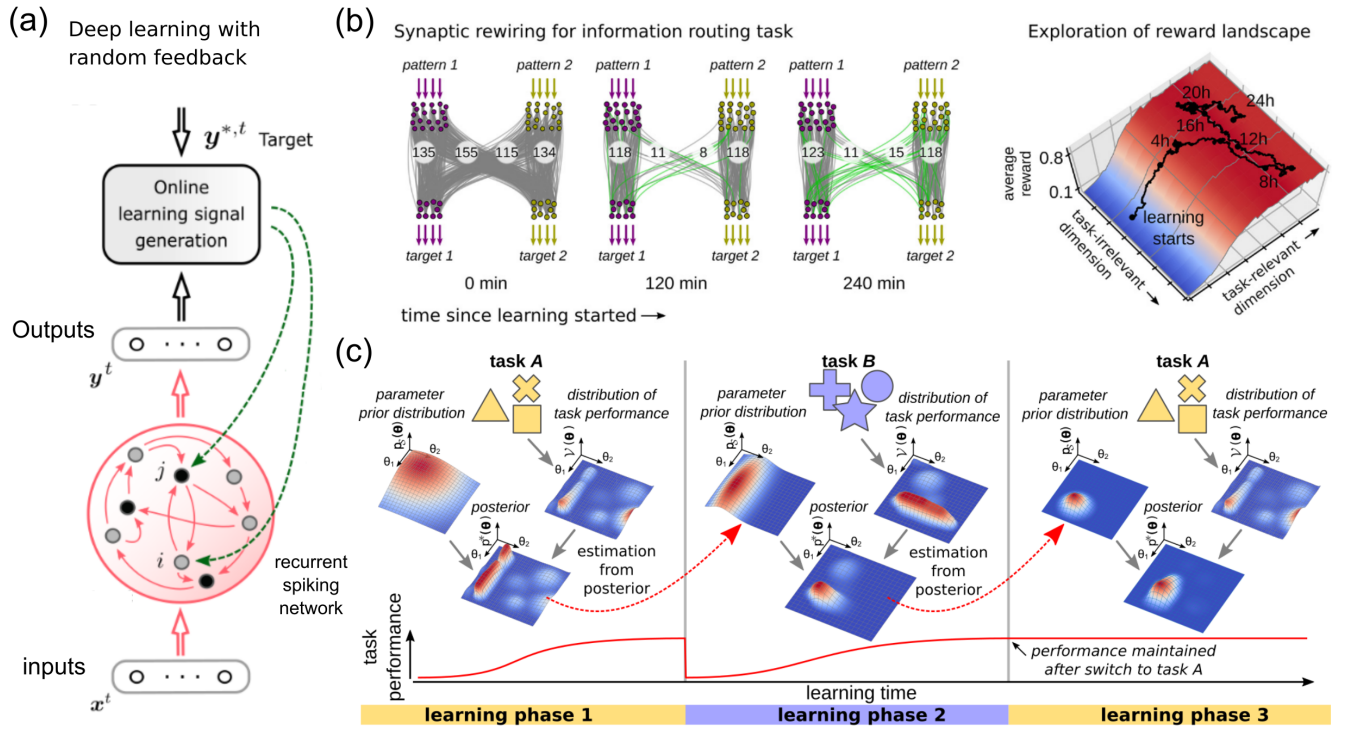
Figure 2. Biologically inspired models of learning in spiking neural networks (a) The e-prop algorithm[95] approximates back-propagation through time using random feedback to propagate error signals to synapses of a recurrent SNN (adapted from[96]) (b) Synaptic sampling[97] exploits the variability of learning rules and redundancy in the task solution space to learn sparse and robust network configurations (adapted from[98]) (c) Overcoming forgetting by selectively slowing down weight changes[99]. After learning a first task A, parameter distributions are absorbed into a prior distribution that confines the motility of synaptic weights in subsequent tasks (task B).

an iterative, gradient-based, supervised learning algorithm that operates in three phases. First, a given input activation is propagated through the network to generate the output based on the current set of parameters. Then, the mismatch between the generated outputs and target values is computed using a loss function, and propagated backwards through the network architecture to compute suitable weight changes. Finally, the network parameters are updated to reduce the loss. We will not go into the details behind Backprop here, but see[1] for an excellent review and historical survey of the development of the algorithm. The problem of porting Backprop to neuromorphic hardware stems form a well-known shortcoming of the algorithm known as *locking* – the weights of a network can only be updated after a full forwards propagation of the data through the network, followed by loss evaluation, then finally after waiting for the back-propagation of error gradients[101]. Locking prevents an efficient implementation of Backprop on online distributed architectures. Also, Backprop is not well suited for spiking neural networks which have non-differentiable output functions. These problems have been recently addressed in brain-inspired variants of the Backprop algorithm.

3.1.1 Brain-inspired alternatives to error backpropagation

In recent years a number of methods have been proposed to approximate the gradient computation performed by Backprop in order to prevent locking (see[102] for a recent review).[103, 104] proposed to replace the non-local error back-propagating term of the Backprop algorithm by sending the loss through a fixed feedback network with random weights that are excluded from training. In this approach, named *random feedback alignment* the back-propagating error signal acts as a local feedback to each synapse, similar to a reward signal in reinforcement learning. The fixed random feedback network de-correlates the error signals providing individual feedback to each synapse. Lillicrap et al. could show that this simple approach already provides a viable approximation to the exact Backprop algorithm and performs well for practical machine learning problems of moderate size. In[105] an event-based version of random feedback alignment, that is well suitable for neuromorphic hardware, was introduced. This approach was further generalized in[106] to include a larger class of algorithms that use error feedback signals.

An efficient model for learning complex sequences in spiking neural networks, named *Superspike*, was introduced in[107]. The model also uses a learning rule that is modulated by error feedback signals and locally minimizes the mismatch between the network output and a target spike train. To overcome the problem of non-differentiable output, Superspike uses a surrogate gradient approach that replaces the infinitely steep spike events with a finite auxiliary function at the time points of network spike events[108, 109]. As in random feedback alignment, learning signals are communicated to the synapses via a feedback network with fixed weights. Using this approach Zenke and others could demonstrate efficient learning of complex sequences in spiking networks.

Another approach to approximate Backprop in spiking neural networks uses an anatomical detail of Cortical neurons.[110] introduced a biologically inspired two-compartment neuron model that approximates the error backpropagation algorithm by minimizing a local dendritic prediction error.[111] port learning by Backprop to neuromorphic hardware by incorporating dynamics with finite time constants and by optimizing the backward pass with respect to substrate variability. They demonstrate the algorithm on the BrainScaleS analog neuromorphic architecture.

3.1.2 Brain-inspired alternatives to backpropagation through time

Recurrent neural network (RNN) architectures often show superior learning results for tasks that involve a temporal dimension, which is often the case for edge computing applications. Porting learning algorithms for RNNs is therefore of utmost importance for efficient machine learning on the edge. Backpropagation through time (BPTT) – the standard RNN learning method used in most GPU implementations – unfolds the network in time and keep this extended structure in memory to propagate information forward and backward which poses a severe challenge to the power and area constraints of edge computing. Recent theoretical results[95, 112] show that the power of BPTT can be brought to biologically inspired spiking neural networks (SNN) while at the same time the unfolding can be prevented in an approximation that operates only forward in time, enabling *online, always-on* learning. This algorithm operates at every synapse in parallel and incrementally updates the synaptic weights. As for random feedback alignment and Superspike discussed above, the weight update depends only on three factors, where the first two are determined by the states of the two related input/output neurons, and the third is given by synapse-specific feedback conveying the mismatch between the target and the actual output (see Fig. 2a for an illustration). The temporal gap between these factors is mitigated by an *eligibility trace* describing a transient dynamic. Eligibility traces, have been theoretically predicted for a long time[113, 114], and have also recently been observed experimentally in the brain[115–118].

3.2 Efficient learning under stringent memory constraints

The amount of available resources in neuromorphic systems is kept low to increase energy efficiency. Memory elements are especially impactful on the energy budget. Therefore, algorithms are needed that make efficient use of the available memory resources. The largest amount of memory in a network is usually consumed by the synaptic weights. Since in practice, the weights of many connections in a network converge to values close to zero, several methods have been proposed to reduce the memory footprint of machine learning algorithms by exploiting sparsity in the network connectivity. We will discuss here two types of algorithms: (1) those that are based on *pruning connections after learning* and (2) *online* learning with *sparse* networks. These two types of sparse learning algorithms are discussed in the following sections.

3.2.1 Pruning

Many approaches to exploit sparsity in learning algorithms focus on pruning the network after training (see[119] for a recent review). Simple methods rely on pruning by magnitude, simply by eliminating the weakest (closest to zero) weights in the network[120–122]. Some methods based on this idea have reported impressive sparsity rates of over 95% for standard machine learning benchmarks with negligible performance loss[123,124]. Other methods are based on theoretical motivations and classical sparsification and regularization techniques[125–127]. These models reach high compression rates.[128] proposed a method to iteratively grow and prune a network in order to generate a compact yet precise solution. They provide a detailed comparison with state of the art dense networks and other pruning methods and reaching sparsity above 99% for the LeNet-5 benchmark.

3.2.2 Online learning in sparse networks

A number of authors also introduced methods that work directly with sparse networks during training, which is often the more interesting case for neuromorphic applications with online training.[129] introduced an algorithm for online stochastic rewiring in deep neural networks that works with a fixed number of synaptic connections throughout learning. The algorithm showed close-to state of the art performance at up to 98% sparsity. Sparse evolutionary training (SET)[130] introduced a heuristic approach that prunes the smallest weights and regrows new weights in random locations. Dynamic Sparse Reparameterization[131] introduces a prune-redistribute-regrowth cycle. They demonstrated compelling performance levels also for very deep neural network architectures.[132] introduced a single shot pruning algorithm that yields sparse networks based on a saliency criterion prior to the actual training.[133] introduced a refined method for online pruning and redistribution that surpasses the previous methods in terms of sparsity and learning performance.

3.3 Open challenges and future work

As outlined above, edge computing poses quite specific challenges to learning algorithms that are substantially different from requirements of classical applications. Some of the algorithms outlined above have already been succesfully ported to neuromorphic hardware. For example, the e-prop algorithm of[112] has been implemented on the SpiNNaker 2 chip yielding an additional energy reduction by two orders of magnitude compared to a X86 implementation[134]. See the next Section 4 for more details on available neuromorphic hardware and their applications.

In the remainder of this section we will highlight open challenges that remain to be solved for efficient learning in edge computing applications. In addition to the stringent memory and power constraints learning at the edge also has to function in an online scenario where data arrive in a continuous stream. Some dedicated hardware resources, e.g. like memristive devices discussed in Section 5, may also show high levels in intrinsic variability, so the learning algorithm should be robust against these noise sources. In this section we discuss recent advances in this line of research and provide food for thought on how these specific challenges can be approached in future work.

3.3.1 Fault-tolerant robust learning algorithms for neuromorphic devices

Here we review recent advances in using inspiration from biology to make learning algorithms robust against device variability. Several authors have suggested that device noise and variability should not be seen as a nuisance, but rather can serve as a computational resource for network simulation and learning algorithms (see[135] for a thorough discussion).[136] have shown that variability in neuronal outputs can be exploited to learn complex statistical dependencies between sensory stimuli. The stochastic behavior of the neurons is used in this model to compute probabilistic inference, while biologically motivated learning rules, that only require local information at the synapses can be used to update the synaptic weights. A theoretical foundation of the model shows that the spiking network performs a Markov chain Monte Carlo sampling process, that allows the network to 'reason' about statistical problems.

This idea is taken one step further in[137] by showing that also the variability of synaptic transmission can be used for stochastic computing. The intrinsic noise of synaptic release is used to drive a sampling process. It was shown that this model can be implemented in an event-based fashion and was benchmarked on the MNIST digit classification task, where it achieved

95.6% accuracy. In[97] it was shown that the variability of learning rules and weight parameters gives rise to a biologically plausible model of online learning. The intrinsic noise of synaptic weight changes drives a sampling process that can be used to exploit redundancies in the task solution space (see Fig. 2b for an illustration). This model was applied to unsupervised learning in spiking neural networks, and to closed-loop reinforcement learning problems[98, 138]. In[139] this model was also ported to the SpiNNaker 2 neuromorphic many-core system.

3.3.2 Biologically motivated mechanisms to combat forgetting in always-on learning scenarios

Neuromorphic systems often operate in an environment where they are permanently on and learning a continuous stream of data. This mode of operation is quite different from most other machine learning applications that work with hand-labeled batches of training data. Always-on learning on a system with limited resources inevitably leads to situations where the system reaches the limits of its memory capacity and thus starts forgetting previously learned sensory experiences. Inspiration to overcome forgetting relevant information comes from biology. The mammalian brain seems to combat forgetting by actively protecting previously acquired knowledge in neocortical circuits[140–144]. When a new skill is acquired, a subset of synapses is strengthened, stabilized and persists despite the subsequent learning of other tasks[143].

A theoretical treatment of the forgetting problem was conducted in the *cascade model* of Stefano Fusi and others[145, 146]. They could show that learning an increasing number of patterns in a single neural network leads unavoidably to a state which they called catastrophic forgetting. Trying to train more patterns into the network will interfere with all previously learned ones, effectively wiping out the information stored in the network. The proposed cascade model to overcome this problem uses multiple parameters per synapse that are linked through a cascade of local interactions. This cascade of parameters selectively slows down weight changes, thus stabilizes synapses when required and effectively combats effects of forgetting. A related model, that uses multiple parameters per synapse to combat forgetting was used in[99] (see also[147] for a recently introduced variation of the model). They used a Bayesian approach that infers a prior distribution over parameter values at each synapse. Synapses that stabilize during learning (converge to a fixed solution) will be considered relevant in subsequent learning and Bayesian priors help to maintain their values (see Fig. 2c for an illustration).

3.3.3 Biologically motivated mechanisms to enhancing transfer and sensor fusion

Distributed computing architectures at the edge need to make decisions by integrate information from different sensors and sensor modalities and they should be able best make use of the sensory information across a wide range of tasks. It is clearly not very efficient to learn from scratch when confronted with a new task. Therefore, to boost the performance of edge computing, we will consider two aspects of transferring information to new situations: transfer of knowledge between sensors (*sensor fusion*), which has been treated in Section 2.2, and transfer of knowledge between multiple different tasks (*transfer learning*).

Transfer learning denotes the improvement of learning in a new task through the use of knowledge from a related task that has already been learned previously[148, 149]. This contrasts most other of today's machine learning applications that focus on one very specific task. In transfer learning, when a new task is learned, knowledge from previous skills can be reused without interfering with them. E.g. the ability to perform a tennis swing can be transferred to playing ping pong, while maintaining the ability to do both sports. The literature on transfer learning is extensive and many different strategies have been developed depending on the relationship between the different task domains (see[150] and[151] for systematic reviews). In machine learning a number of approaches have been applied to a wide range of problems, including classification of images[152–155], text[156–159] or human activity[160].

A very general approach to learn across multiple domains is followed in the *learning to learn* framework of[161, 162]. Their model features networks that are able to modify their own weights through the network activity. These network are therefore able to tinker with their own processing properties. This approach has been taken to its most extreme form where a network leans to implement an optimization algorithm by itself[163]. This model consists of an outer-loop learning network (*the optimizer*) that controls the parameters of an inner-loop network (*the optimizee*). The training algorithm of the inner-loop network works on single tasks that are presented sequentially, whereas the outer-loop learner operates across tasks and can acquire strategies to transfer knowledge. This learning-to-learn framework was recently applied to SNNs to obtain properties of LSTM networks and use them to solve complex sequence learning tasks[112]. In[164] the learning-to-learn framework was also applied to a neuromorphic hardware platform.

4 Signal processing for wearable devices on neuromorphic chip

Neuromorphic engineering is a branch of electrical engineering dedicated to the design of analog/digital data processors that aims to emulate biological neurons and synapses. It typically consumes less energy than conventional computing systems and presents additional properties, such as massively parallel event-based computation, distributed local memory and

adaptation[165, 166]. This increasing interest in neuromorphic engineering shows that hardware SNNs are considered a key future technology with high potential in key application, such as the Edge of Computing, and wearable devices.

Neuromorphic technologies have sparked interest from universities[8, 12, 167–169] and companies such as IBM[9] and Intel[10]. In this Section, we will provide an overview of the neuromorphic platforms, that to the best of our knowledge were deployed for biomedical signal processing, showing promising results to be exploited in wearable devices.

4.1 Neuromorphic processors

TrueNorth. TrueNorth[9] is IBM's fully digital neuromorphic chip with one million neurons arranged in a tiled array of 4096 neurosynaptic cores enabling *massive parallel processing*. Each core contains 13kB of *local SRAM memory* to keep neurons and synapse's states along with the axonal delays and information on the fan-out destination. There are 256 Leaky-Integrator and Fire (LIF) neurons implemented by time-multiplexing and 256 million synapses are designed in the form of SRAM memory. Each core can support up to 256 fan-in and fan-out, and this connectivity can be configured such that a neuron in any core can communicate its spikes any other neuron in any other core.

Thanks to the *event-driven*, the co-location of memory and processing units in each core, and the use of low-leakage silicon CMOS technology, TrueNorth can perform 46 billion synaptic operations per second (SOPS) per watt for real-time operation, with 26 pJ per synaptic event. Its power density of 20 mW/cm^2 is about three orders of magnitude smaller than that of typical CPUs.

SpiNNaker. The SpiNNaker machine[8], designed by the University of Manchester, is a custom-designed ASIC based on *massively parallel architecture* that has been designed to efficiently simulate large spiking neural networks. It consists of ARM968 processing cores arranged in a 2D array where the precise details of the neurons and their dynamics can be programmed into. Although the processing cores are synchronous microprocessors, the *event-based* aspect of SpiNNaker is apparent in its message-handling paradigm. A message (event) gets delivered to a core generating a request for being processed. The communications infrastructure between these nodes is specially optimized to carry very large numbers of very small packets, optimal for spiking neurons.

A second generation of SpiNNaker was designed by Technical University of Dresden[170]. Spinnaker2 continues the line of dedicated digital neuromorphic chips for brain simulation increasing the simulation capacity by a factor > 10 while staying in the same power budget (i.e. 10x better power efficiency). The full-scale SpiNNaker2 consists of 10 Million ARM cores distributed across 70000 Chips in 10 server racks. This system takes advantage of advanced 22nm FDSOI technology node with Adaptive Body Biasing enabling reliable and ultra-low power processing. It also features incorporating numerical accelerators for the most common operations.

Loihi. Loihi[10] is Intel's neuromorphic chip with many core processing incorporating on-line learning designed in 14 nm FinFET technology. The chip supports about 130000 neurons and 130 million synapses distributed in 128 cores. Spikes are transported between the cores in the chip using packetized messages by an asynchronous network on chip. It includes three embedded x86 processors and provides a very flexible learning engine on which diverse online learning algorithms such as Spike-Timing Dependent Plasticity (STDP), different 3 factor and trace-based learning rules can be implemented. The chip also provides hierarchical connectivity, dendritic compartments, synaptic delays as different features that can enrich a spiking neural network. The synaptic weights are stored on local SRAM memory and the bit precision can vary between 1 to 9 bits. All logic in the chip is digital, functionally deterministic, and implemented in an asynchronous bundled data design style.

DYNAP-SE. DYNAP-SE implements a multi-core neuromorphic processor with scalable architecture fabricated using a standard 0.18 μm CMOS technology[12]. It is a full-custom asynchronous mixed-signal processor, with a fully asynchronous inter-core and inter-chip hierarchical routing architecture. Each core comprises 256 adaptive exponential integrate-and-fire (AEI&F) neurons for a total of 1k neurons per chip. Each neuron has a Content Addressable Memory (CAM) block, containing 64 addresses representing the pre-synaptic neurons that the neuron is subscribed to. Rich synaptic dynamics are implemented on the chip by using Differential Pair Integrator (DPI) circuits [171]. These circuits produce EPSCs and IPSCs (Excitatory/Inhibitory Post Synaptic Currents), with time constants that can range from a few μs to hundreds of *ms*. The analog circuits are operated in the sub-threshold domain, thus minimizing the dynamic power consumption, and enabling implementations of neural and synaptic behaviors with biologically plausible temporal dynamics. The asynchronous CAMs on the synapses are used to store the tags of the source neuron addresses connected to them, while the SRAM cells are used to program the address of the destination core/chip that the neuron targets.

ODIN/MorphIC. ODIN (Online-learning DIgital spiking Neuromorphic) processor occupies an area of only 0.086mm^2 in 28nm FDSOI CMOS[13]. It consists of a single neurosynaptic core with 256 neurons and 256^2 synapses. Each neuron can be configured to phenomenologically reproduce the 20 Izhikevich behaviors of spiking neurons[172]. The synapses embed a 3-bit weight and a mapping table bit that allows enabling or disabling Spike-Dependent Synaptic Plasticity (SDSP) locally[173], thus allowing for the exploration of both off-chip training and on-chip online learning setups.

MorphIC is a quad-core digital neuromorphic processor with 2k LIF neurons and more than 2M synapses in 65nm CMOS[174].

Table 1. Summary of neuromorphic platforms and biomedical applications

Neuromorphic Chip	DYNAP-SE	SpiNNaker	Loihi	TrueNorth	ODIN
CMOS Technology	180nm	ARM968, 130 nm	14nm FinFET	28nm	28 nm FDSOI
Implementation	Mixed-signal	Digital	Digital ASIC	Digital ASIC	Digital ASIC
Energy per SOP	17 pJ @ 1.8V	Peak power 1W per chip	23.6 pJ @ 0.75V	26 pJ @ 0.775	12.7 pJ@0.55V
Size	38.5 mm^2	102 mm^2	60 mm^2	0.093 mm^2 (core)	0.086 mm^2
On-chip learning	No	Yes (configurable)	Yes (configurable)	No	Yes (SDSP)
Applications	EMG, ECG, HFO	EMG and EEG	EMG	EEG and Local Field Potential (LFP)	EMG

MorphIC was designed for high-density large-scale integration of multi-chip setups. The four 512-neuron crossbar cores are connected with a hierarchical routing infrastructure that enables neuron fan-in and fan-out values of 1k and 2k, respectively. The synapses are binary and can be either programmed with offline-trained weights or trained online with a stochastic version of SDSP.

4.2 Biomedical signal processing on Neuromorphic hardware

Table 1 shows the summary of neuromorphic processors described previously and in which biomedical signal processing applications were used. These works show promising results for always-on embedded biomedical systems.

The first chip presented in this table is DYNAP-SE, used to implement SNNs for the classification or detection of EMG[175, 176] and ECG[177, 178] and to implement a simple spiking perceptron as part of a design to detect High Frequency Oscillation (HFO) in human intracranial EEG[179]. In particular, in[175, 177] a spiking RNN is deployed for ECG/EMG signal separation to facilitate the classification with a linear read-out. SVM and linear least square approximation is used in the read out layer for[177, 178] and overall accuracy of 91% and 95% for anomaly detection were reached respectively. In[175], the state property of the spiking RNN on EMG was investigated for different hand gestures. In[176] the performance of a feedforward SNN and a hardware-friendly spiking learning algorithm for hand gesture recognition using superficial EMG was investigated and compared to traditional machine learning approaches, such as SVM. Results show that applying SVM on the spiking output of the hidden layer achieved a classification rate of 84%, and the spiking learning method achieved 74% with a power consumption of about 0.05 mW. The consumption was compared to state-of-the-art embedded system showing that the proposed spiking network is two orders of magnitude more power efficient[180, 181].

Recently, the benchmark hand-gesture classification was processed and compared on two other digital neuromorphic platforms, i.e. Loihi and ODIN/MorphIC[13, 174]. A spiking Convolutional Neural Network (CNN) was implemented on Loihi and a spiking Multilayer Perceptron (MLP) was implemented on ODIN/MorphIC[182]. Because of the properties of neuromorphic chips, on Loihi a late fusion was implemented combining the output from the spiking CNN for vision, and the spiking MLP for EMG signals; While on ODIN/MorphIC hardware, the two spiking MLPs were fused in the last layer. Due to the neuromorphic chip properties the Loihi implemented a late fusion of a spiking CNN, for vision and a spiking MLP for EMG signals. In the ODIN/MorphIC system two spiking MLPs were fused in the last layer. The comparison with the embedded GPU was performed in terms of accuracy, power consumption, and latency showing that the neuromorphic chips are able to achieve the same accuracy with significantly smaller energy-delay product, 30x and 600x more efficient for Loihi and ODIN/MorphIC, respectively[182].

4.3 Encoding

In SNNs a single spike by itself does not carry any information. However, the number and the timing of spikes produced by a neuron are important. Just as their biological counterpart, silicon neurons in neuromorphic devices produce spike trains at a rate that is proportional to their input current. At the input side, synapse circuits integrate the spikes they receive to produce analog currents, with temporal dynamics and time constants that can be made equivalent to their biological counterparts. The sum of all the positive (excitatory) and negative (inhibitory) synaptic currents afferent to the neuron is then injected into the neuron.

To provide biomedical signals to the synapses of the SNN input layer, it is necessary to first convert them into spikes. A common way to do this is to use a delta-modulator circuit[179, 183] functionally equivalent to the one used in the Dynamic Vision Sensor (DVS)[184]. This circuit, in practice, is an ADC that produces two asynchronous digital pulse outputs (UP or DOWN) for

every biosignal channel in the input. The UP (DOWN) spikes are generated every time the difference between the current and previous value exceeds a pre-defined threshold. The sign of the difference corresponds to the UP or DOWN channel where the spike is produced. This approach was used to convert EMG signals, used in mixed-signal neuromorphic chips[175, 176] and in digital ones[182, 185], ECG signals[177, 178], and EEG and HFO ones[179, 183].

4.4 Adaptation in neuromorphic processor

Local adaptation is an important aspect in extreme edge computing, specially when it comes to wearable devices. The current methods for training networks for biomedical signals rely on large datasets collected from different patients. However, when it comes to biological data, there is no "one size fits all". Each patient and person has their own unique biological signature. Therefore, the field of Personalized Medicine (PM) has gained lots of attention in the past few years and the online on-edge adaptation feature of neuromorphic chips can be a game changer for PM.

As was discussed in Section 3.1, there are lots of effort in designing spike-based online learning algorithms which can be implemented on neuromorphic chips.

Example of today's state of the art for on-chip learning are Intel's Loihi[10], DynapSEL and ROLLS chip from UZH/ETHZ[168, 186], BrainScales from Heidelberg[11] and ODIN from UC Louvain[13]. Intel's Loihi includes a learning engine which can implement different learning rules such as simple pairwise STDP, triplet STDP, reinforcement learning with synaptic tag assignments or any 3 factor learning rule implementation. DynapSEL, ROLLS and ODIN encompass the SDSP, also known as the Fusi learning rule, which is a form of semi-supervised learning rule that can support both unsupervised clustering applications and supervised learning with labels for shallow networks[173]. BrainscaleS chip implements the STDP rule. Moreover, Spinnaker 1 and 2[170, 187] can implement a wide variety of on-chip learning algorithms since their designs make use of ARM microcontrollers providing lots of configurability for the users.

4.5 Open challenges

Generally, implementing on-chip online learning is challenging because of these two core reasons: locality of the weight update and weight storage.

Locality The learning information for updating the weights of any on-chip network should be locally available to the synapse since otherwise this information should be "routed" to the synapse by wires which will take a significant amount of area on chip. The simplest form of learning which satisfies this requirement is Hebbian learning which has been implemented on a variety of neuromorphic chips forms of unsupervised/semi-supervised learning[11, 13, 168, 186]. However, Hebbian-based algorithms are limited in the tasks they can learn and to the best of our knowledge no large scale task has been demonstrated using this rule. Since gradient descent-based algorithms such as Backprop has had lots of success in deep learning, there are more and more spike-based error Backprop rules that are being developed as was discussed in Section 3.1. These types of learning algorithms have recently been custom designed in the form of spike-based delta rule as back-bone of the Backprop algorithm. For example, single layer implementation of the delta rule has been designed in[188] and employed for EMG classification[176]. Expanding this to multi-layer networks involves non-local weight updates which limits its on-chip implementation. Making the Backprop algorithm local is a topic of on-going research which we have discussed in Section 3.1. Recently, a multi-layer perceptron error-triggered learning architecture has been proposed to overcome the non-locality of multi-layer networks solving the spatial credit assignment problem on chip[106, 189]

Weight storage The ideal weight storage for online on-chip learning should have the following properties: (i) non-volatility to keep the state of the learnt weights even when the power shuts down to reduce the time and energy footprints of reloading the weights to the chip. (ii) Linear update which allows the state of the memory to change linearly with the calculated update. (iii) Analog states which allows a full-precision for the weights. Non-volatile memristive devices have been proposed as a great potential for the weight storage and there is a large body of work combining the CMOS technology with that of the memristive devices to get the best of two worlds.

In the next Section we provide a thorough review on the state of the art for the emerging memory devices and the efforts to integrate and use them in conjunction with neuromorphic chips.

5 Memristive devices and computing

The severe power and area constraints under which a neuromorphic processor for edge computing must work opened ways towards the investigation of beyond-CMOS solutions. Despite still at the dawn of its technological development, memristive devices have been drawing attention in the last decade thanks to their scalability, low-power operation, compatibility with CMOS chip power supply and CMOS fabrication process, and volatile/non-volatile properties. In Section 5.1, we will introduce memristive devices and the properties that are appealing for adaptive extreme edge computing paradigms. In Section 5.2,
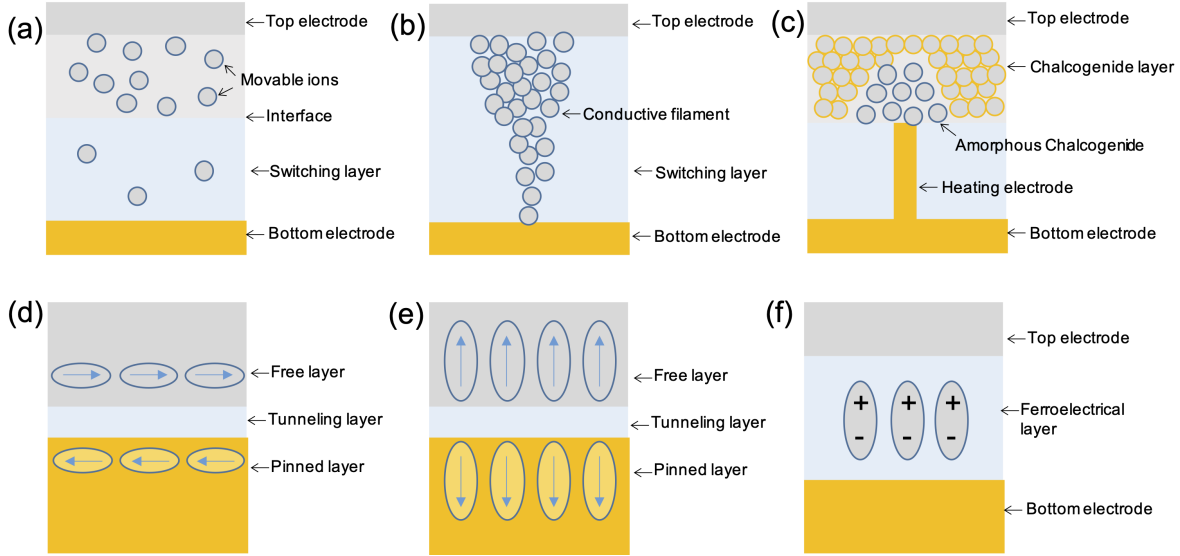
Figure 3. Memristive devices for neuromorphic computing. (a) Interface type RRAM device; (b) Filamentary RRAM device; (c) Phase change memory device; (d) MRAM device with in-plane spin polarization; (e) MRAM device with perpendicular spin polarization; (f) FTJ device.

we will explore the role of memristive devices in neuromemristive systems and give examples of possible applications. In Section 5.3, we will discuss the current challenges and the future perspectives of memristive technology.

5.1 Conventional and wearable memristive devices

Memristive devices, as the name suggested, are devices which can change and memorize their resistance states. They are usually two-terminal devices, however, can be implemented with various physical mechanisms, resulting in versatile existing forms, e.g. resistive random access memory (RRAM, Fig. 3a and 3b) ([25]), phase change memory (PCM, Fig. 3c) ([190]), magnetic random access memory (MRAM, Fig. 3d and Fig. 3e) ([191]), ferroelectric tunneling junction (FTJ, Fig. 3f) ([192]), etc. The resistance memory of these devices can mimic the memory effect of the basic components of biological neural system, while the resistance changing can mimic the plasticity of biological synapse. Facilitated with their simplicity of two-terminal configuration and scalability to nanoscale, they are inherently suitable for the hardware implementation of brain-inspired computation materializing an artificial neural network, i.e. neuromorphic computation ([193, 194]).

This notation, in recent years, has incited wide investigations on the various memristive devices and on their applications in neural network learning and recognition, or, in short, memristive learning ([195–200]). The memristive learning can enable energy efficient and low latency information process within a reduced size of systems abandoning the conventional von-Neumann architecture. Among other benefits, this will also make it possible to process information where they are acquired, i.e. within sensors, and reduce the bandwidth needed for transferring the sensor data to data center, accelerating the coming of the era of Internet-of-Things (IOT). Table 2 summarizes the key features of the main memristive device technologies for neuromorphic / wearable applications in terms of cell area, electrical characteristics, main advantages and challenges. It is worth noticing that some figures of merit in this context are radically different with respect to standard memory requirements. Indeed, while in the memory scenario higher read currents enable faster reading speed, in neuromorphic applications currents as low as possible are preferred, since the current is a limiting factor for neurons' fan-out. Similarly, SET and RESET times should be as fast as possible in memory applications, while in our applications this requirement can be relaxed thanks to the lower operating frequency of the neurons (20 Hz to 100 Hz). Moreover, the number achievable conductance levels has to be increased ([201]). Some non-idealities which are usually detrimental for memory applications, for instance stochasticity of switching parameters, are even beneficial for the neural networks.

In addition to the commonly referred non-volatile type of memristive switching, the RRAM device can also show volatile behavior, which usually occurs when active materials such as silver or copper are used as electrode. The relatively long retention time of the volatile behavior (tens of milliseconds to seconds) is then found to be similar to the timescale of short term memory, and naturally was proposed to mimic the short term memory effect of biological synapses ([20, 23, 218]).

Although most researches on memristive devices are carried on rigid silicon substrates, the simple structure of memristive devices can also be realized on flexible substrates ([219]), which opens new interesting possibilities for realizing local computation

Table 2. Key features of non-volatile memristive devices.

	RRAM	PCM	MRAM	FTJ
Cell area [min. feature size]	$4F^2$ [202]	$4F^2$ [202]	$9F^2$ ([203])	$4F^2$ [202]
Retention	>10 years ([16])	>10 years ([14])	>10 years ([17])	>10 years ([15])
Endurance	10^{12} ([204,205])	10^{11} ([206])	10^{12} ([207])	$> 10^{15}$ ([15])
SET / RESET time	100 ps ([208]) 85 ps ([210])	>100 ns, 10 ns ([202])	20 ns ([209]) 3 ns ([211])	30 ns, 30 ns ([212])
Read current	100 pA ([213])	25 μA ([214])	20 μA ([211])	0.8 nA ([215], device diameter 300 nm)
Write energy per bit	20 fJ ([216])	~100 fJ ([217])	90 fJ ([211])	<10 fJ ([212])
Main features	Scalability, speed, low energy	Scalability, multilevel, low voltage	Endurance, low power	Endurance, low power, speed
Challenges	Variability	RESET current, temperature stability, resistance drift	Density, scalability, variability	Scalability

within wearable devices ([220,221]).

5.2 Memristive devices for neuromorphic computing

5.2.1 Memristive neural components

As mentioned in Section 5.1, the primary function of memristive devices is the usage as synaptic devices to implement the memory and plasticity of biological synapses. However, there are increasing interests for these devices to be utilized to implement nanoscale and artificial neurons.

On the neuron side, the memristive device gradual internal state change and its consequently abrupt switching closely mimic the integrate-and-fire behavior of biological neurons ([222,224,225], Fig. 4a-c). Due to the sample structure and nanometer level scalability, memristive neurons can be much more compact than current CMOS neurons which might consist of current sensor, analog-to-digital converter (ADC), and analog-to-digital converter (DAC), and capacitors, all of which are expensive to implement in current CMOS technology in terms of area and/or power consumption ([226]). The implementation of memristive neurons will also enable full memristive neuromorphic computing ([227]), which promises further increases in the integration of the hardware neuromorphic computing.

On the synaptic side, the key feature of the biological synapses is their plasticity, i.e. tunable weight, which can be generally implemented by resistance or conductance modification in the memristive devices (Fig. 4d). Fundamental learning rules based on STDP have already been widely explored ([196,228–231]). Spatial spiking pattern recognition ([232]), spiking co-incidence detection ([233,234]), and spatial-temporal correlation ([223,235]) has been reported recently. Synaptic metaplasticity, such as paired-pulse facilitation, can also be achieved via various device operation mechanism ([20,236,237]).

5.2.2 Memristive neural network architectures

There are generally two approaches for a hardware neuromorphic system implementing memristive devices as synapses: (i) deep learning accelerator, accelerating the artificial neural network computing with multiple layer and error back-propagation, as well as it's variations, like convolutional neural network, recurrent neural network, etc.; (ii) brain-like computing, attempting to closely mimicking the behaviors of biological neural system, like spike representation (Fig. 4d) and collective decision making behavior. In the deep learning accelerator approach, on-line training places more requirements for the memristive synapses. For instance, linear and symmetrical weight update is crucial for the on-line training ([200,238]), while off-line training ignores it since the synaptic weight can be programmed to the memristive device with fine tuning and iterative verify ([239]).

Collective decision making is an important feature of the brain computing, which requires high parallelism and, consequently, low current devices. For instance, this feature is the essential for Hopfield neural network ([240]), cellular neural network ([241]), and coupled oscillators ([242]). In the Hopfield neural network, the system automatically evolves to its energy minimization points leading the functionality of associative memory. The use of Hopfield like recurrent neural networks (RNNs) with memristive devices has already been successfully demonstrated in a variety of tasks ([243,244]). As an example of memristive based coupled oscillator network,[245] used a network of self-sustained van der Pol oscillators coupled with oxide-based memristive devices to
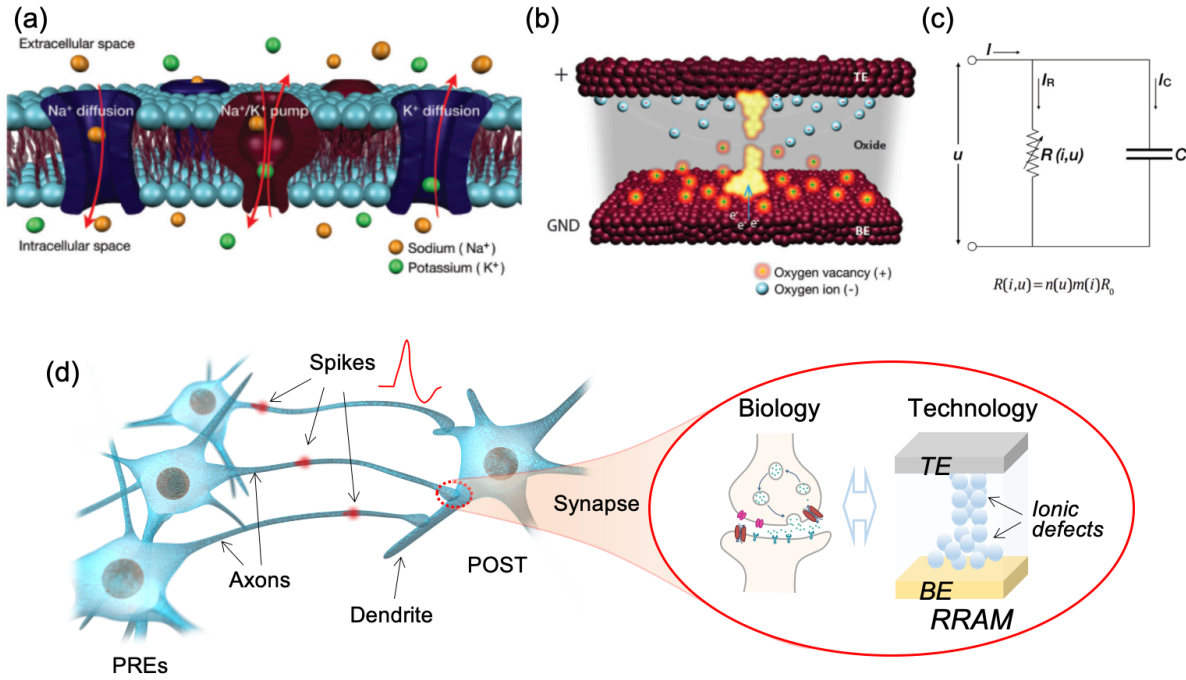
Figure 4. Memristive devices as synapse or neuron for neuromorphic computing. (a)-(c) memristive device act as threshold device for the firing function of biological neuron ([222], reproduced under the CC BY license). (d) Conceptual illustration of memristive device as artificial synapse for brain-like neuromorphic computing ([223], reproduced under the CC BY-NC license).

investigate the temporal binding problem, which is a well known issue in the field of cognitive neuroscience. In this experiment, the network is able to emulate an optical illusion which shows two patterns depending on the influence of attention. This means that the network is able to select relevant information from a pool of inputs, as in the case of a system collecting signals from multiple sensors.

5.2.3 Applications of memristive neural networks

At present, memristive technology has been mainly used in relatively simple networks with Hebbian-based learning algorithms. However, more recently, systems able of solving different tasks, such as speech recognition ([246]), and exploring different architectures and learning algorithms are being investigated. In particular, the benefits of exploiting sparsity, mentioned in Section 3.2, are demonstrated for feature extraction and image classification in networks trained with stochastic gradient descend and winner-take-all learning algorithms ([247]), as well as in hierarchical temporal memory, which does not need training ([248]).

In the latest years, memristive devices have been used in applications closer to biology, enabling hybrid biological-artificial systems ([249]) and investigating biomedical applications, ranging from speech and emotion recognition ([250]) to biosignal ([251]) and medical image ([252]) processing. Finally, an interesting application is the one of memristive biosensors, which[253] used to implement a system for cancer diagnostic. The innovative use of memristive properties was demonstrated in hardware and opens the way to a broader use of memristive technology where sensors and computing co-exist in the same system or, possibly, in the same device.

5.3 Open challenges and future work

5.3.1 Device non-idealities

Implementation of mainstream deep learning algorithms with Backprop learning rule and memristive synapses imposes some requirements for the memristive device, including linear current-voltage relation for reading, analog conductance tuning, linear and symmetric weight update, long retention time, high endurance, etc. ([254]). However, no single device can fulfill all these requirements simultaneously.

Various techniques have been proposed to compensate the device non-idealities. For instance, to compensate the non-linear current-voltage relation for reading, fixed read voltage with variable pulse width or pulse number can be used for synaptic

weight reading, and the readout is represented by the charge accumulation in the output nodes ([255]). Linear and symmetric weight update is crucial for accurate online learning of a memristive multilayer neural network with Backprop learning rule ([238]). However, PCM devices usually only show gradual switching in set direction (weight potentiation), while RRAM devices show gradual switching in reset direction (weight depression). To achieve linear and symmetric weight update, differential pair with two of these devices are usually used. For a differential pair with two PCM devices, the potentiation is achieved by applying set pulses on the positive part and the depression is achieved by applying set pulses on the negative part, thus gradual weight update in both potentiation and depression can be achieved. To further enhance the linearity of weight update, a minor conductance pair consisting of capacitors can be used for frequent but smaller weight update, and finally transferred to the major pair periodically ([200]). Another option to improve device linearity is limiting the device dynamic range in a region far from saturation and where the weight update is linear[256, 257].

In addition to mitigate the non-idealities of memristive devices, more and more research efforts are made to exploit these non-idealities for brain-like computations. For instance, the stochasticity or noise in reading of memristive device can be used for the probability computation for restricted Boltzmann machine ([258]), or escape for local minimization points in a Hopfield neural network ([259]). The Ag filament based resistive switching device shows short retention time and high switching dynamics, thus was proposed for reservoir computing ([260]) and spatiotemporal computing ([218]) to process time-encoded information.

5.3.2 Co-integration of hybrid CMOS-memristive neuromorphic systems

The main steps to be taken to exploit the full potential of an ASIC for end-to-end processing system go through the integration of memristive devices and sensors with CMOS technology. Indeed, the works presented so far are based either on simulations or on real device data, or on memristive chips interfaced with some standard digital hardware. Despite integration of CMOS technology has been demonstrated for non-volatile resistive switching devices already at a commercial level ([261, 262]), the design of co-integrated memristive-based neuromorphic processors is still under development. We envisage a three-phase process to achieve a fully integrated system.

The first one is the co-integration of non-volatile memristive devices with some peripheral circuits ([263]) and to implement some logic and multiply-and-accumulate (MAC) operations ([264]), which reaches the maturity with the demonstration of a fully cointegrated SNN with analog neurons and memristive synapses ([265]). The second phase is the co-integration of different technologies. Despite this approach results in higher fabrication costs, it presents several advantages in terms of system performance, which can be more compact and potentially more power efficient. In particular, the co-integration of non-volatile and volatile memristive devices can lead to a fully memristive approach. As an example,[227] exploit volatile memristive devices to emulate stochastic neurons and non-volatile memristive devices to store the synaptic weights on the same chip, thus demonstrating the feasibility and the advantages of the dual technology co-integration process. Eventually, the final step which has to be taken in the development of a dedicated ASIC for wearable edge computing is the co-integration of sensors and memristive-based systems.[266] tackled this challenge by designing and fabricating a gas sensing system able of gas classification. The system uses RRAM arrays as memory, Carbon Nanotube field effect transistor (CNFET) for computation and gas sensing, both 3D monolithically integrated on CMOS circuits, which carry out computation and allow memory access.

5.3.3 Learning with memristive devices

Adaptability is a feature of paramount importance in smart wearable devices, which need to be able to learn the unique feature of their user. This calls for the implementation of lifelong learning paradigms, i.e. the ability of continuously learning new features from experience. Typically, a network has a limited memory capacity dependent on the network size and architecture. Once the maximum number of experiences is recorded, new features learned will erase old ones, thus originating the phenomenon of catastrophic forgetting.

The problem of an efficient implementation of continual learning has been thoroughly investigated ([267]). In the current scenario, a dichotomy exist between backprop-based ANNs, which have very high accuracy but a limited memory capacity, and brain-inspired SNNs, which feature higher memory capacity thanks to their higher flexibility, but at the cost of lower accuracy. Models used to overcome forgetting are described in Section 3.3. The use of memristive devices in such networks is still an open point. It is possible that memristive device will be beneficial to increase the network capacity ([268]) at no extra computational cost thanks to their slow approach to the boundaries ([269]), but so far this topic is still quite unexplored. An interesting approach is proposed by[270], where the key strengths of supervised convolutional ANNs, unsupervised SNNs, and memristive devices are combined in a single system. The results indicate that this approach is robust against catastrophic forgetting, whilst reaching 93% accuracy when tested with both trained and non-trained classes.

6 Discussion and Conclusions

In this study, we presented the state-of-the-art core elements that enable the development of wearable devices with extreme edge adaptive computing capability. Various sensors that can collect different bio-signals from the human body are investigated.

There is a variety of sensing specifications in terms of size, resolution, mechanical flexibility and output signals need to be considered along with their analogue readout circuit at a limited amount of power consumption. However, when the real-time processing of these signals is deployed on edge, severe constraints raise in terms of power efficiency, fast response times, and accuracy in the data classification. The widely-used solution is to find a trade-off between the energy and computational capacity, or send the data to the cloud. However, these strategies are not ideal and slow down the development of wearable smart sensing. To meet all the requirements, the development of a platform needs to be optimized in synergy with the other elements and every aspect of the design, from the learning algorithms to the architecture.

In particular, continual learning is required for adaptive wearable devices. In this respect, brain-inspired algorithms promise to be valid alternatives to standard machine learning approaches such as Backprop and BPTT. The exploitation of sparsity in network connectivity increases the power efficiency by optimizing the use of the available memory. However, the problem of algorithmic robustness to non-ideal hardware (such as noise and variability) and the problems of forgetting and information transfer between tasks still persist and have to be solved in combination with neuromorphic and emerging technologies. SNNs are conceptually ideal for low-power in-memory computing. Their event-based approach together with the use of analog subthreshold circuits to reproduce biological timescales, allows fast response times of the network while enabling smooth real-time processing of data. The encoding of the incoming signals into spikes is however still challenging. Moreover, a fully CMOS-based approach has two major technological issues. First, the synaptic weight is usually stored in SRAMs, which hold the state only in the presence of a power supply. Second, capacitors used to implement biological time constants are massive and may consume up to 60% of the chip area. Memristive technology can be beneficial in this respect. Non-volatile devices can potentially replace SRAMs and volatile devices offer a compact alternative to CMOS capacitors. Besides low-power operation in a small footprint, memristive devices also offer noisy properties, which – if exploited in the right way – might facilitate the implementation of stochastic learning algorithms. However, the technology is still at its infancy and fabrication processes are still under development, yielding high device variability, which makes it difficult to produce reliable multi-bit memory.

In summary, the ultimate goal towards smart wearable sensing with edge computing capabilities relies on a bespoke platform consists of embedding sensors, front-end circuit interface, neuromorphic processor and memristive devices. This platform requires high-compatibility of existing sensing technologies with CMOS circuitry and memristive devices to move the intelligent algorithm into the wearable edge without significantly increase the cost in energy. New solutions are needed to enhance the performance of local adaptive learning rules to be competitive with the accuracy of Backprop. Novel encoding techniques to allow streamless communication from sensors to neuromorphic chip have to be developed and flanked by efficient event-based algorithms. So far there is not a uniquely ideal solution, but we envisage that a holistic approach where all the elements of the system are co-designed as a whole is the key to build low-power end-to-end real-time adaptive systems for next-generation smart wearable devices.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

All the Authors equally contributed to the manuscript, actively participating to the discussions and to the writing. The main contributors for each Section are as follows: X.L. and H.H. – wearable sensors; D.K. – biologically plausible models; M.P. and E.D. – signal processing and neuromorphic computing. E.C. and W.W. – memristive devices. E.C. led and coordinated the cooperative writing and all discussions.

Funding

Acknowledgments

References

1. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks* **61**, 85–117 (2015).

2. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444, DOI: 10.1038/nature14539 (2015).

3. Silver, D. *et al.* Mastering the game of go with deep neural networks and tree search. *nature* **529**, 484–489 (2016).

4. Chen, Y.-H., Krishna, T., Emer, J. S. & Sze, V. Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks. *IEEE J. Solid-state Circuits* **52**, 127–138 (2016).

5. Cavigelli, L. & Benini, L. Origami: A 803-GOp/s/W Convolutional Network Accelerator. *IEEE Transactions on Circuits Syst. for Video Technol.* **27**, 2461–2475 (2016).

6. Song, J. *et al.* An 11.5TOPS/W 1024-MAC Butterfly Structure Dual-Core Sparsity-Aware Neural Processing Unit in 8nm Flagship Mobile SoC. In *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC)*, 130–132 (San Francisco, CA., 2019).

7. Lee, J. *et al.* LNPU: A 25.3 TFLOPS/W Sparse Deep-Neural-Network Learning Processor with Fine-Grained Mixed Precision of FP8-FP16. In *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC)*, 142–144 (San Francisco, CA., 2019).

8. Furber, S. B., Galluppi, F., Temple, S. & Plana, L. A. The spinnaker project. *Proc. IEEE* **102**, 652–665 (2014).

9. Merolla, P. A. *et al.* A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).

10. Davies, M. *et al.* Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).

11. Schemmel, J. *et al.* A wafer-scale neuromorphic hardware system for large-scale neural modeling. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, 1947–1950 (2010).

12. Moradi, S., Qiao, N., Stefanini, F. & Indiveri, G. A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (dynaps). *IEEE transactions on biomedical circuits systems* **12**, 106–122 (2017).

13. Frenkel, C., Lefebvre, M., Legat, J.-D. & Bol, D. A 0.086-mm^2 12.7-pj/sop 64k-synapse 256-neuron online-learning digital spiking neuromorphic processor in 28-nm cmos. *IEEE Transactions on Biomed. Circuits Syst.* **13**, 145–158 (2019).

14. Cheng, H. Y. *et al.* A thermally robust phase change memory by engineering the ge/n concentration in (ge, n)$_x$sb$_y$te$_z$ phase change material. In *2012 International Electron Devices Meeting*, 31.1.1–31.1.4 (2012).

15. Udayakumar, K. R. *et al.* Low-power ferroelectric random access memory embedded in 180nm analog friendly cmos technology. In *2013 5th IEEE International Memory Workshop*, 128–131 (2013).

16. Goux, L. *et al.* Role of the Ta scavenger electrode in the excellent switching control and reliability of a scalable low-current operated TiN / Ta$_2$O$_5$ / Ta RRAM device. In *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 1–2 (2014).

17. Golonzka, O. *et al.* Mram as embedded non-volatile memory solution for 22ffl finfet technology. In *2018 IEEE International Electron Devices Meeting (IEDM)*, 18.1.1–18.1.4 (2018).

18. Jo, S. H., Kumar, T., Narayanan, S. & Nazarian, H. Cross-Point Resistive RAM Based on Field-Assisted Superlinear Threshold Selector. *IEEE Transactions on Electron Devices* **62**, 3477–3481, DOI: 10.1109/TED.2015.2426717 (2015).

19. Yang, H. *et al.* Threshold switching selector and 1S1R integration development for 3D cross-point STT-MRAM. In *2017 IEEE International Electron Devices Meeting (IEDM)*, 38.1.1–38.1.4, DOI: 10.1109/IEDM.2017.8268513 (2017).

20. Wang, Z. *et al.* Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat. Mater.* **16**, 101–108, DOI: 10.1038/nmat4756 (2017).

21. Wang, W. *et al.* Surface diffusion-limited lifetime of silver and copper nanofilaments in resistive switching devices. *Nat. Commun.* **10**, 81, DOI: 10.1038/s41467-018-07979-0 (2019).

22. Wang, W., Covi, E., Lin, Y., Ambrosi, E. & Ielmini, D. Modeling of switching speed and retention time in volatile resistive switching memory by ionic drift and diffusion. In *2019 IEEE International Electron Devices Meeting (IEDM)*, 32.3.1–32.3.4, DOI: 10.1109/IEDM19573.2019.8993625 (2019).

23. Covi, E. *et al.* A volatile rram synapse for neuromorphic computing. In *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 903–906, DOI: 10.1109/ICECS46596.2019.8965044 (2019).

24. Linares-Barranco, B. & Serrano-Gotarredona, T. Memristance can explain spike-time-dependent-plasticity in neural synapses. *Nat. Preced.* (2009).

25. Ielmini, D. & Wong, H.-S. P. In-memory computing with resistive switching devices. *Nat. Electron.* **1**, 333–343, DOI: 10.1038/s41928-018-0092-2 (2018). 1801.06601.

26. Chicca, E. & Indiveri, G. A recipe for creating ideal hybrid memristive-cmos neuromorphic processing systems. *Appl. Phys. Lett.* **116**, 120501, DOI: 10.1063/1.5142089 (2020). https://doi.org/10.1063/1.5142089.

27. Gao, W. *et al.* Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature* **529**, 509–514, DOI: 10.1038/nature16521 (2016).

28. Kanoun, O. & Tränkler, H. R. Sensor technology advances and future trends. *IEEE Transactions on Instrumentation Meas.* **53**, 1497–1501, DOI: 10.1109/TIM.2004.834613 (2004).

29. López, A., Fernández, M., Rodríguez, H., Ferrero, F. & Postolache, O. Development of an eog-based system to control a serious game. *Meas. J. Int. Meas. Confed.* **127**, 481–488, DOI: 10.1016/j.measurement.2018.06.017 (2018).

30. Nweke, H. F., Teh, Y. W., Al-garadi, M. A. & Alo, U. R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert. Syst. with Appl.* **105**, 233–261, DOI: 10.1016/j.eswa.2018.03.056 (2018).

31. Witkowski, M. *et al.* Enhancing brain-machine interface (bmi) control of a hand exoskeleton using electrooculography (eog). *J. NeuroEngineering Rehabil.* **11**, 1–6, DOI: 10.1186/1743-0003-11-165 (2014).

32. Herry, C. L., Frasch, M., Seely, A. J. E. & Wu, H. T. Heart beat classification from single-lead ecg using the synchrosqueezing transform. *Physiol. Meas.* **38**, 171–187, DOI: 10.1088/1361-6579/aa5070 (2017).

33. Pantelopoulos, A. & Bourbakis, N. G. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Syst. Man, Cybern. Part C (Applications Rev.* **40**, 1–12, DOI: 10.1109/TSMCC.2009.2032660 (2010).

34. Li, H., Shrestha, A., Heidari, H., Le Kernec, J. & Fioranelli, F. A multisensory approach for remote health monitoring of older people. *IEEE J. Electromagn. RF Microwaves Med. Biol.* **2**, 102–108 (2018).

35. Liang, X. *et al.* Fusion of wearable and contactless sensors for intelligent gesture recognition. *Adv. Intell. Syst.* **1**, 1900088, DOI: 10.1002/aisy.201900088 (2019).

36. He, S., Yang, C., Wang, M., Cheng, L. & Hu, Z. Hand gesture recognition using myo armband. *Proc. - 2017 Chin. Autom. Congr. CAC 2017* **2017-Janua**, 4850–4855, DOI: 10.1109/CAC.2017.8243637 (2017).

37. Khezri, M. & Jahed, M. Real-time intelligent pattern recognition algorithm for surface emg signals. *BioMedical Eng. Online* **6**, 1–12, DOI: 10.1186/1475-925X-6-45 (2007).

38. Liang, X., Ghannam, R. & Heidari, H. Wrist-worn gesture sensing with wearable intelligence. *IEEE Sensors J.* DOI: 10.1109/JSEN.2018.2880194 (2018).

39. Wu, W., Nagarajan, S. & Chen, Z. Bayesian machine learning: Eegmeg signal processing measurements. *IEEE Signal Process. Mag.* **33**, 14–36, DOI: 10.1109/MSP.2015.2481559 (2016).

40. Yazicioglu, R. F., Van Hoof, C. & Puers, R. *Biopotential readout circuits for portable acquisition systems* (Springer Science & Business Media, 2008).

41. Luz, E. J. d. S., Schwartz, W. R., Cámara-Chávez, G. & Menotti, D. Ecg-based heartbeat classification for arrhythmia detection: A survey. *Comput. Methods Programs Biomed.* **127**, 144–164, DOI: https://doi.org/10.1016/j.cmpb.2015.12.008 (2016).

42. Kiranyaz, S., Ince, T. & Gabbouj, M. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomed. Eng.* **63**, 664–675, DOI: 10.1109/TBME.2015.2468589 (2016).

43. Rahhal, M. M. A. *et al.* Deep learning approach for active classification of electrocardiogram signals. *Inf. Sci.* **345**, 340–354, DOI: 10.1016/j.ins.2016.01.082 (2016).

44. Raj, S., Ray, K. C. & Shankar, O. Cardiac arrhythmia beat classification using dost and pso tuned svm. *Comput. Methods Programs Biomed.* **136**, 163–177, DOI: 10.1016/j.cmpb.2016.08.016 (2016).

45. Zhang, Z., Dong, J., Luo, X., Choi, K.-S. & Wu, X. Heartbeat classification using disease-specific feature selection. *Comput. Biol. Medicine* **46**, 79–89, DOI: https://doi.org/10.1016/j.compbiomed.2013.11.019 (2014).

46. Alfaras, M., Soriano, M. C. & Ortín, S. A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. *Front. Phys.* **7**, DOI: 10.3389/fphy.2019.00103 (2019).

47. Ortín, S., Soriano, M. C., Alfaras, M. & Mirasso, C. R. Automated real-time method for ventricular heartbeat classification. *Comput. Methods Programs Biomed.* **169**, 1–8, DOI: 10.1016/J.CMPB.2018.11.005 (2019).

48. Hossain, M. S. & Muhammad, G. Cloud-assisted industrial internet of things (iiot) – enabled framework for health monitoring. *Comput. Networks* **101**, 192–202, DOI: https://doi.org/10.1016/j.comnet.2016.01.009 (2016).

49. Yang, Z., Zhou, Q., Lei, L., Zheng, K. & Xiang, W. An iot-cloud based wearable ecg monitoring system for smart healthcare. *J. Med. Syst.* **40**, 286, DOI: 10.1007/s10916-016-0644-9 (2016).

50. Jebelli, H., Hwang, S. & Lee, S. Eeg signal-processing framework to obtain high-quality brain waves from an off-the-shelf wearable eeg device. *J. Comput. Civ. Eng.* **32**, 04017070, DOI: 10.1061/(ASCE)CP.1943-5487.0000719 (2018).

51. Lin, C. *et al.* Wireless and wearable eeg system for evaluating driver vigilance. *IEEE Transactions on Biomed. Circuits Syst.* **8**, 165–176, DOI: 10.1109/TBCAS.2014.2316224 (2014).

52. Gargiulo, G. *et al.* A new eeg recording system for passive dry electrodes. *Clin. Neurophysiol.* **121**, 686–693, DOI: https://doi.org/10.1016/j.clinph.2009.12.025 (2010).

53. Thakor, N. V. Biopotentials and electrophysiology measurements. In *Telehealth and Mobile Health*, 595–614 (CRC press, 2015).

54. Li, G., Lee, B. & Chung, W. Smartwatch-based wearable eeg system for driver drowsiness detection. *IEEE Sensors J.* **15**, 7169–7180, DOI: 10.1109/JSEN.2015.2473679 (2015).

55. Shen, K.-Q., Li, X.-P., Ong, C.-J., Shao, S.-Y. & Wilder-Smith, E. P. V. Eeg-based mental fatigue measurement using multi-class support vector machines with confidence estimate. *Clin. Neurophysiol.* **119**, 1524–1533, DOI: https://doi.org/10.1016/j.clinph.2008.03.012 (2008).

56. Wang, X.-W., Nie, D. & Lu, B.-L. Emotional state classification from eeg data using machine learning approach. *Neurocomputing* **129**, 94–106, DOI: https://doi.org/10.1016/j.neucom.2013.06.046 (2014).

57. BioSemi. Biosemi systems. https://www.biosemi.com/products.htm (2020).

58. Hosseinifard, B., Moradi, M. H. & Rostami, R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal. *Comput. Methods Programs Biomed.* **109**, 339–345, DOI: 10.1016/j.cmpb.2012.10.008 (2013).

59. Hwang, S., Jebelli, H., Choi, B., Choi, M. & Lee, S. Measuring workers' emotional state during construction tasks using wearable eeg. *J. Constr. Eng. Manag.* **144**, 04018050, DOI: 10.1061/(ASCE)CO.1943-7862.0001506 (2018).

60. Xu, J., Mitra, S., Hoof, C. V., Yazicioglu, R. F. & Makinwa, K. A. A. Active electrodes for wearable eeg acquisition: Review and electronics design methodology. *IEEE Rev. Biomed. Eng.* **10**, 187–198, DOI: 10.1109/RBME.2017.2656388 (2017).

61. Duchowski, A. *Eye Tracking Methodology - Theory and Practice* (Springer, Cham, 2007).

62. Eid, M. A., Giakoumidis, N. & Saddik, A. E. A novel eye-gaze-controlled wheelchair system for navigating unknown environments: Case study with a person with als. *IEEE Access* **4**, 558–573, DOI: 10.1109/ACCESS.2016.2520093 (2016).

63. Duvinage, M., Castermans, T. & Dutoit, T. Control of a lower limb active prosthesis with eye movement sequences. In *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, 1–7, DOI: 10.1109/CCMB.2011.5952116 (2011).

64. Barua, S., Ahmed, M. U., Ahlström, C. & Begum, S. Automatic driver sleepiness detection using eeg, eog and contextual information. *Expert. Syst. with Appl.* **115**, 121–135, DOI: https://doi.org/10.1016/j.eswa.2018.07.054 (2019).

65. Piñero, P. *et al.* Sleep stage classification using fuzzy sets and machine learning techniques. *Neurocomputing* **58-60**, 1137–1143, DOI: https://doi.org/10.1016/j.neucom.2004.01.178 (2004).

66. Zhu, X. *et al.* Eog-based drowsiness detection using convolutional neural networks. In *2014 International Joint Conference on Neural Networks (IJCNN)*, 128–134, DOI: 10.1109/IJCNN.2014.6889642 (2014).

67. Martin, W. B. *et al.* Pattern recognition of eeg-eog as a technique for all-night sleep stage scoring. *Electroencephalogr. Clin. Neurophysiol.* **32**, 417–427, DOI: https://doi.org/10.1016/0013-4694(72)90009-0 (1972).

68. Stevens, J. R. *et al.* Telemetered eeg-eog during psychotic behaviors of schizophrenia. *Arch. Gen. Psychiatry* **36**, 251–262, DOI: 10.1001/archpsyc.1979.01780030017001 (1979).

69. Punsawad, Y., Wongsawat, Y. & Parnichkun, M. Hybrid eeg-eog brain-computer interface system for practical machine control. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 1360–1363, DOI: 10.1109/IEMBS.2010.5626745 (2010).

70. Wang, H., Li, Y., Long, J., Yu, T. & Gu, Z. An asynchronous wheelchair control by hybrid eeg–eog brain–computer interface. *Cogn. Neurodynamics* **8**, 399–409, DOI: 10.1007/s11571-014-9296-y (2014).

71. Mendez, I. *et al.* Evaluation of the myo armband for the classification of hand motions. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, 1211–1214, DOI: 10.1109/ICORR.2017.8009414 (2017).

72. Rissanen, S. M. *et al.* Surface emg and acceleration signals in parkinson's disease: feature extraction and cluster analysis. *Med. & Biol. Eng. & Comput.* **46**, 849–858, DOI: 10.1007/s11517-008-0369-0 (2008).

73. Wang, Q. *et al.* A novel pedestrian dead reckoning algorithm using wearable emg sensors to measure walking strides. In *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, 1–8, DOI: 10.1109/UPINLBS.2010.5653821 (2010).

74. Rawat, S., Vats, S. & Kumar, P. Evaluating and exploring the myo armband. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, 115–120, DOI: 10.1109/SYSMART.2016.7894501 (2016).

75. Inhyuk, M., Myungjoon, L., Junuk, C. & Museong, M. Wearable emg-based hci for electric-powered wheelchair users with motor disabilities. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2649–2654, DOI: 10.1109/ROBOT.2005.1570513 (2005).

76. Artemiadis, P. K. & Kyriakopoulos, K. J. A switching regime model for the emg-based control of a robot arm. *IEEE Transactions on Syst. Man, Cybern. Part B (Cybernetics)* **41**, 53–63, DOI: 10.1109/TSMCB.2010.2045120 (2011).

77. Cipriani, C., Zaccone, F., Micera, S. & Carrozza, M. C. On the shared control of an emg-controlled prosthetic hand: Analysis of user–prosthesis interaction. *IEEE Transactions on Robotics* **24**, 170–184, DOI: 10.1109/TRO.2007.910708 (2008).

78. Subasi, A. Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders. *Comput. Biol. Medicine* **43**, 576–586, DOI: https://doi.org/10.1016/j.compbiomed.2013.01.020 (2013).

79. Rincon, A. L., Yamasaki, H. & Shimoda, S. Design of a video game for rehabilitation using motion capture, emg analysis and virtual reality. In *2016 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 198–204, DOI: 10.1109/CONIELECOMP.2016.7438575 (2016).

80. Biswas, D., Simões-Capela, N., Hoof, C. V. & Helleputte, N. V. Heart rate estimation from wrist-worn photoplethysmography: A review. *IEEE Sensors J.* **19**, 6560–6570, DOI: 10.1109/JSEN.2019.2914166 (2019).

81. Biswas, D. *et al.* Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment. *IEEE Transactions on Biomed. Circuits Syst.* **13**, 282–291, DOI: 10.1109/TBCAS.2019.2892297 (2019).

82. Reşit Kavsaoğlu, A., Polat, K. & Recep Bozkurt, M. A novel feature ranking algorithm for biometric recognition with ppg signals. *Comput. Biol. Medicine* **49**, 1–14, DOI: https://doi.org/10.1016/j.compbiomed.2014.03.005 (2014).

83. Caytak, H., Boyle, A., Adler, A. & Bolic, M. Bioimpedance spectroscopy processing and applications. In Narayan, R. (ed.) *Encyclopedia of Biomedical Engineering*, 265 – 279, DOI: https://doi.org/10.1016/B978-0-12-801238-3.10884-0 (Elsevier, Oxford, 2019).

84. Matthie, J. R. Bioimpedance measurements of human body composition: critical analysis and outlook. *Expert. Rev. Med. Devices* **5**, 239–261, DOI: 10.1586/17434440.5.2.239 (2008).

85. Sun, T.-P. *et al.* The use of bioimpedance in the detection/screening of tongue cancer. *Cancer Epidemiol.* **34**, 207–211, DOI: https://doi.org/10.1016/j.canep.2009.12.017 (2010).

86. Zhang, Y., Xiao, R. & Harrison, C. Advancing hand gesture recognition with high resolution electrical impedance tomography. *UIST 2016 - Proc. 29th Annu. Symp. on User Interface Softw. Technol.* 843–850, DOI: 10.1145/2984511.2984574 (2016).

87. Alsheikh, M. A., Lin, S., Niyato, D. & Tan, H. P. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Commun. Surv. Tutorials* **16**, 1996–2018, DOI: 10.1109/COMST.2014.2320099 (2014).

88. Gravina, R., Alinia, P., Ghasemzadeh, H. & Fortino, G. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Inf. Fusion* **35**, 1339–1351, DOI: 10.1016/j.inffus.2016.09.005 (2017).

89. Khaleghi, B., Khamis, A., Karray, F. O. & Razavi, S. N. Multisensor data fusion: A review of the state-of-the-art. *Inf. Fusion* **14**, 28–44 (2013).

90. Rundo, F., Conoci, S., Ortis, A. & Battiato, S. An advanced bio-inspired photoplethysmography (ppg) and ecg pattern recognition system for medical assessment. *Sensors (Basel)* **18**, DOI: 10.3390/s18020405 (2018).

91. He, X., Goubran, R. A. & Liu, X. P. Secondary peak detection of ppg signal for continuous cuffless arterial blood pressure measurement. *IEEE Transactions on Instrumentation Meas.* **63**, 1431–1439, DOI: 10.1109/TIM.2014.2299524 (2014).

92. Chiu, H.-Y., Shuai, H.-H. & Chao, P. C.-P. Reconstructing qrs complex from ppg by transformed attentional neural networks. *IEEE SENSORS JOURNAL* (2020).

93. Patel, S. *et al.* A wearable computing platform for developing cloud-based machine learning models for health monitoring applications. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5997–6001, DOI: 10.1109/EMBC.2016.7592095 (2016).

94. Mead, C. How we created neuromorphic engineering. *Nat. Electron.* **3**, 434–435, DOI: 10.1038/s41928-020-0448-2 (2020).

95. Bellec, G., Scherr, F., Hajek, E. *et al.* Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. *arXiv preprint arXiv:1901.09049* (2019).

96. Bellec, G. *et al.* A solution to the learning dilemma for recurrent networks of spiking neurons. *bioRxiv* 738385 (2020).

97. Kappel, D., Habenschuss, S., Legenstein, R. & Maass, W. Network plasticity as bayesian inference. *PLoS computational biology* **11** (2015).

98. Kappel, D., Legenstein, R., Habenschuss, S., Hsieh, M. & Maass, W. A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning. *Eneuro* **5** (2018).

99. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J. *et al.* Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* 201611835 (2017).

100. Rumelhart, D., Hinton, G. & Williams, R. Learning internal representations by error propagation. In *In: Parallel Distributed Processing*, vol. 1, 318–362 (MIT Press, Cambridge, MA, 1986).

101. Czarnecki, W. M. *et al.* Understanding synthetic gradients and decoupled neural interfaces. *arXiv preprint arXiv:1703.00522* (2017).

102. Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nat. neuroscience* **22**, 1761–1770 (2019).

103. Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random synaptic feedback weights support error backpropagation for deep learning. *Nat. communications* **7**, 1–10 (2016).

104. Samadi, A., Lillicrap, T. P. & Tweed, D. B. Deep learning with dynamic spiking neurons and fixed feedback weights. *Neural computation* **29**, 578–602 (2017).

105. Neftci, E. O., Augustine, C., Paul, S. & Detorakis, G. Event-driven random back-propagation: Enabling neuromorphic deep learning machines. *Front. neuroscience* **11**, 324 (2017).

106. Payvand, M., Fouda, M. E., Kurdahi, F., Eltawil, A. & Neftci, E. O. Error-triggered three-factor learning dynamics for crossbar arrays. In *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 218–222 (IEEE, 2020).

107. Zenke, F. & Ganguli, S. Superspike: Supervised learning in multilayer spiking neural networks. *Neural computation* **30**, 1514–1541 (2018).

108. Hinton, G., Srivastava, N. & Swersky, K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on* **14** (2012).

109. Bengio, Y., Léonard, N. & Courville, A. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432* (2013).

110. Sacramento, J., Costa, R. P., Bengio, Y. & Senn, W. Dendritic error backpropagation in deep cortical microcircuits. *arXiv preprint arXiv:1801.00062* (2017).

111. Göltz, J. *et al.* Fast and deep neuromorphic learning with time-to-first-spike coding. *arXiv preprint arXiv:1912.11443* (2019).

112. Bellec, G., Salaj, D., Subramoney, A., Legenstein, R. & Maass, W. Long short-term memory and learning-to-learn in networks of spiking neurons. In *Advances in Neural Information Processing Systems*, 787–797 (2018).

113. Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. learning* **8**, 229–256 (1992).

114. Izhikevich, E. M. Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cereb. cortex* **17**, 2443–2452 (2007).

115. Yagishita, S. *et al.* A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616–1620 (2014).

116. He, K., Huertas, M., Hong, S. Z. *et al.* Distinct eligibility traces for ltp and ltd in cortical synapses. *Neuron* **88**, 528–538 (2015).

117. Brzosko, Z., Schultz, W. & Paulsen, O. Retroactive modulation of spike timing-dependent plasticity by dopamine. *Elife* **4**, e09685 (2015).

118. Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S. & Magee, J. C. Behavioral time scale synaptic plasticity underlies ca1 place fields. *Science* **357**, 1033–1036 (2017).

119. Gale, T., Elsen, E. & Hooker, S. The state of sparsity in deep neural networks. *arXiv preprint arXiv:1902.09574* (2019).

120. Ström, N. Sparse connection and pruning in large dynamic artificial neural networks. In *5th European Conference on Speech Communication and Technology*, 2807–2810 (1997).

121. Collins, M. D. & Kohli, P. Memory bounded deep convolutional networks. *arXiv preprint arXiv:1412.1442* (2014).

122. Han, S., Pool, J., Tran, J. & Dally, W. Learning both weights and connections for efficient neural network. In *Advances in neural information processing systems*, 1135–1143 (2015).

123. Guo, Y., Yao, A. & Chen, Y. Dynamic network surgery for efficient dnns. In *Advances in neural information processing systems*, 1379–1387 (2016).

124. Zhu, M. & Gupta, S. To prune, or not to prune: exploring the efficacy of pruning for model compression. *arXiv preprint arXiv:1710.01878* (2017).

125. Molchanov, D., Ashukha, A. & Vetrov, D. Variational dropout sparsifies deep neural networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2498–2507 (JMLR. org, 2017).

126. Louizos, C., Welling, M. & Kingma, D. P. Learning sparse neural networks through l_0 regularization. *arXiv preprint arXiv:1712.01312* (2017).

127. Ullrich, K., Meeds, E. & Welling, M. Soft weight-sharing for neural network compression. *arXiv preprint arXiv:1702.04008* (2017).

128. Dai, X., Yin, H. & Jha, N. K. Nest: A neural network synthesis tool based on a grow-and-prune paradigm. *IEEE Transactions on Comput.* **68**, 1487–1497 (2019).

129. Bellec, G., Kappel, D., Maass, W. & Legenstein, R. Deep rewiring: Training very sparse deep networks. *arXiv preprint arXiv:1711.05136* (2017).

130. Mocanu, D. C. *et al.* Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nat. communications* **9**, 1–12 (2018).

131. Mostafa, H. & Wang, X. Parameter efficient training of deep convolutional neural networks by dynamic sparse reparameterization. *arXiv preprint arXiv:1902.05967* (2019).

132. Lee, N., Ajanthan, T. & Torr, P. H. Snip: Single-shot network pruning based on connection sensitivity. *arXiv preprint arXiv:1810.02340* (2018).

133. Dettmers, T. & Zettlemoyer, L. Sparse networks from scratch: Faster training without losing performance. *arXiv preprint arXiv:1907.04840* (2019).

134. Liu, C. *et al.* Memory-efficient deep learning on a spinnaker 2 prototype. *Front. neuroscience* **12**, 840 (2018).

135. Maass, W. Noise as a resource for computation and learning in networks of spiking neurons. *Proc. IEEE* **102**, 860–880 (2014).

136. Pecevski, D. & Maass, W. Learning probabilistic inference through spike-timing-dependent plasticity. *eneuro* **3** (2016).

137. Neftci, E. O., Pedroni, B. U., Joshi, S., Al-Shedivat, M. & Cauwenberghs, G. Unsupervised learning in synaptic sampling machines. *arXiv preprint arXiv:1511.04484* (2015).

138. Kaiser, J. *et al.* Embodied synaptic plasticity with online reinforcement learning. *Front. Neurorobotics* **13**, 81 (2019).

139. Yan, Y. *et al.* Efficient reward-based structural plasticity on a spinnaker 2 prototype. *IEEE transactions on biomedical circuits systems* **13**, 579–591 (2019).

140. Cichon, J. & Gan, W.-B. Branch-specific dendritic Ca2+ spikes cause persistent synaptic plasticity. *Nature* **520**, 180 (2015).

141. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowl. Data Eng.* **22**, 1345–1359 (2009).

142. Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F. *et al.* Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* **525**, 333 (2015).

143. Yang, G., Pan, F. & Gan, W.-B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).

144. Yang, G. *et al.* Sleep promotes branch-specific formation of dendritic spines after learning. *Science* **344**, 1173–1178 (2014).

145. Fusi, S., Drew, P. J. & Abbott, L. F. Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).

146. Benna, M. K. & Fusi, S. Computational principles of synaptic memory consolidation. *Nat. neuroscience* **19**, 1697–1706 (2016).

147. Huszár, F. Note on the quadratic penalties in elastic weight consolidation. *Proc. Natl. Acad. Sci.* 201717042 (2018).

148. Caruana, R. Multitask learning. *Mach. learning* **28**, 41–75 (1997).

149. Torrey, L. & Shavlik, J. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, 242–264 (IGI Global, 2010).

150. Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *J. Big Data* **3**, 9 (2016).

151. Lu, J. *et al.* Transfer learning using computational intelligence: a survey. *Knowledge-Based Syst.* **80**, 14–23 (2015).

152. Long, M., Zhu, H., Wang, J. & Jordan, M. I. Deep transfer learning with joint adaptation networks. In *34th International Conference on Machine Learning-Volume 70*, 2208–2217 (JMLR. org, 2017).

153. Duan, L., Xu, D. & Tsang, I. Learning with augmented features for heterogeneous domain adaptation. *arXiv preprint arXiv:1206.4660* (2012).

154. Kulis, B., Saenko, K. & Darrell, T. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *CVPR 2011*, 1785–1792 (IEEE, 2011).

155. Zhu, Y. *et al.* Heterogeneous transfer learning for image classification. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 1304–1309 (2011).

156. Wang, C. & Mahadevan, S. Heterogeneous domain adaptation using manifold alignment. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 1541–1546 (2011).

157. Zhou, J. T., Tsang, I. W., Pan, S. J. & Tan, M. Heterogeneous domain adaptation for multiple classes. In *Artificial Intelligence and Statistics*, 1095–1103 (2014).

158. Prettenhofer, P. & Stein, B. Cross-language text classification using structural correspondence learning. In *48th Annual Meeting of the Association for Computational Linguistics*, 1118–1127 (2010).

159. Zhou, J. T., Pan, S. J., Tsang, I. W. & Yan, Y. Hybrid heterogeneous transfer learning through deep learning. In *Twenty-eighth AAAI conference on artificial intelligence*, 2213–2219 (2014).

160. Harel, M. & Mannor, S. Learning from multiple outlooks. *arXiv preprint arXiv:1005.0027* (2010).

161. Schmidhuber, J. Learning to control fast-weight memories: An alternative to dynamic recurrent networks. *Neural Comput.* **4**, 131–139 (1992).

162. Schmidhuber, J. A neural network that embeds its own meta-levels. In *IEEE International Conference on Neural Networks*, 407–412 (IEEE, 1993).

163. Andrychowicz, M. *et al.* Learning to learn by gradient descent by gradient descent. In *Advances in neural information processing systems*, 3981–3989 (2016).

164. Bohnstingl, T., Scherr, F., Pehle, C., Meier, K. & Maass, W. Neuromorphic hardware learns to learn. *Front. neuroscience* **13**, 483 (2019).

165. Indiveri, G. & Liu, S.-C. Memory and information processing in neuromorphic systems. *Proc. IEEE* **103**, 1379–1397 (2015).

166. Chicca, E., Stefanini, F., Bartolozzi, C. & Indiveri, G. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proc. IEEE* **102**, 1367–1388 (2014).

167. Schemmel, J., Billaudelle, S., Dauer, P. & Weis, J. Accelerated analog neuromorphic computing. *arXiv preprint arXiv:2003.11996* (2020).

168. Qiao, N. *et al.* A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses. *Front. neuroscience* **9**, 141 (2015).

169. Neckar, A. *et al.* Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model. *Proc. IEEE* **107**, 144–164 (2018).

170. Mayr, C., Hoeppner, S. & Furber, S. Spinnaker 2: A 10 million core processor system for brain simulation and machine learning. *arXiv preprint arXiv:1911.02385* (2019).

171. Bartolozzi, C. & Indiveri, G. Synaptic dynamics in analog vlsi. *Neural computation* **19**, 2581–2603 (2007).

172. Izhikevich, E. M. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks* **15**, 1063–1070 (2004).

173. Brader, J. M., Senn, W. & Fusi, S. Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural computation* **19**, 2881–2912 (2007).

174. Frenkel, C., Legat, J.-D. & Bol, D. Morphic: A 65-nm 738k-synapse/mm^2 quad-core binary-weight digital neuromorphic processor with stochastic spike-driven online learning. *IEEE Transactions on Biomed. Circuits Syst.* **13**, 999–1010 (2019).

175. Donati, E. *et al.* Processing emg signals using reservoir computing on an event-based neuromorphic system. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–4 (IEEE, 2018).

176. Donati, E., Payvand, M., Risi, N., Krause, R. B. & Indiveri, G. Discrimination of emg signals using a neuromorphic implementation of a spiking neural network. *IEEE transactions on biomedical circuits systems* (2019).

177. Bauer, F. C., Muir, D. R. & Indiveri, G. Real-time ultra-low power ecg anomaly detection using an event-driven neuromorphic processor. *IEEE transactions on biomedical circuits systems* (2019).

178. Corradi, F. *et al.* Ecg-based heartbeat classification in neuromorphic hardware. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE, 2019).

179. Sharifshazileh, M., Burelo, K., Fedele, T., Sarnthein, J. & Indiveri, G. A neuromorphic device for detecting high-frequency oscillations in human ieeg. In *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 69–72 (IEEE, 2019).

180. Benatti, S. *et al.* A Versatile Embedded Platform for EMG Acquisition and Gesture Recognition. *IEEE Transactions on Biomed. Circuits Syst.* **9**, 620–630 (2015).

181. Montagna, F., Rahimi, A., Benatti, S., Rossi, D. & Benini, L. PULP-HD: Accelerating Brain-inspired High-dimensional Computing on a Parallel Ultra-low Power Platform. In *Proceedings of the ACM/ESDA/IEEE Design Automation Conference (DAC)*, 1–6 (San Francisco, CA., 2018).

182. Ceolini, E. *et al.* Hand-gesture recognition based on emg and event-based camera sensor fusion: a benchmark in neuromorphic computing. *Front. Neurosci.* 36 (2020).

183. Corradi, F. & Indiveri, G. A neuromorphic event-based neural recording system for smart brain-machine-interfaces. *IEEE transactions on biomedical circuits systems* **9**, 699–709 (2015).

184. Lichtsteiner, P., Posch, C. & Delbruck, T. A 128x128 120 db 15 us latency asynchronous temporal contrast vision sensor. *IEEE journal solid-state circuits* **43**, 566–576 (2008).

185. Behrenbeck, J. *et al.* Classification and regression of spatio-temporal signals using neucube and its realization on spinnaker neuromorphic hardware. *J. neural engineering* **16**, 026014 (2019).

186. Qiao, N. & Indiveri, G. Scaling mixed-signal neuromorphic processors to 28 nm FD-SOI technologies. In *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 552–555 (IEEE, Shanghai, China., 2016).

187. Furber, S. B. *et al.* Overview of the SpiNNaker System Architecture. *IEEE Transactions on Comput.* **62**, 2454–2467, DOI: http://doi.ieeecomputersociety.org/10.1109/TC.2012.142 (2013).

188. Payvand, M. & Indiveri, G. Spike-based Plasticity Circuits for Always-on On-line Learning in Neuromorphic Systems. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5 (Sapporo, Japan., 2019).

189. Payvand, M., Fouda, M., Kurdahi, F., Eltawil, A. & Neftci, E. On-chip error-triggered learning of multi-layer memristive spiking neural networks. *J. Emerg. Technol. Circuits Syst. (JETCAS)* (2020).

190. Zhang, W., Mazzarello, R., Wuttig, M. & Ma, E. Designing crystallization in phase-change materials for universal memory and neuro-inspired computing. *Nat. Rev. Mater.* **4**, 150–168, DOI: 10.1038/s41578-018-0076-x (2019).

191. Miron, I. M. *et al.* Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. *Nature* **476**, 189–193, DOI: 10.1038/nature10309 (2011).

192. Wen, Z., Li, C., Wu, D., Li, A. & Ming, N. Ferroelectric-field-effect-enhanced electroresistance in metal/ferroelectric/semiconductor tunnel junctions. *Nat. Mater.* **12**, 617–621, DOI: 10.1038/nmat3649 (2013).

193. Jo, S. H. *et al.* Nanoscale Memristor Device as Synapse in Neuromorphic Systems. *Nano Lett.* **10**, 1297–1301, DOI: 10.1021/nl904092h (2010).

194. Wang, W. *et al.* A hardware neural network for handwritten digits recognition using binary RRAM as synaptic weight element. In *2016 IEEE Silicon Nanoelectronics Workshop (SNW)*, 50–51 (IEEE, 2016).

195. Ohno, T. *et al.* Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591–595, DOI: 10.1038/nmat3054 (2011).

196. Kuzum, D., Jeyasingh, R. G. D., Yu, S. & Wong, H.-S. P. Low-Energy Robust Neuromorphic Computation Using Synaptic Devices. *IEEE Transactions on Electron Devices* **59**, 3489–3494, DOI: 10.1109/TED.2012.2217146 (2012).

197. Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nat. Nanotechnol.* **8**, 13–24, DOI: 10.1038/nnano.2012.240 (2013).

198. Alibart, F., Zamanidoost, E. & Strukov, D. B. Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nat. Commun.* **4**, 2072, DOI: 10.1038/ncomms3072 (2013).

199. Eryilmaz, S. B. *et al.* Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front. Neurosci.* **8**, 1–11, DOI: 10.3389/fnins.2014.00205 (2014). 1406.4951.

200. Ambrogio, S. *et al.* Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67, DOI: 10.1038/s41586-018-0180-5 (2018).

201. Ielmini, D. & Pedretti, G. Device and circuit architectures for in-memory computing. *Adv. Intell. Syst.* **n/a**, 2000040 (2020).

202. IRDS. International Roadmap for Devices and Systems™. https://irds.ieee.org/ (2020).

203. Rho, K. *et al.* 23.5 a 4gb lpddr2 stt-mram with compact 9f2 1t1mtj cell and hierarchical bitline architecture. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 396–397 (2017).

204. Kim, Y. *et al.* Bi-layered rram with unlimited endurance and extremely uniform switching. In *2011 Symposium on VLSI Technology - Digest of Technical Papers*, 52–53 (2011).

205. Lee, M.-J. *et al.* A fast, high-endurance and scalable non-volatile memory device made from asymmetric ta_2o_{5-x}/tao_{2-x} bilayer structures. *Nat. Mater.* **10**, 625–630, DOI: 10.1038/nmat3070 (2011).

206. Kim, I. S. *et al.* High performance pram cell scalable to sub-20nm technology with below $4f^2$ cell size, extendable to dram applications. In *2010 Symposium on VLSI Technology*, 203–204 (2010).

207. Saida, D. *et al.* $1\times$ - to $2\times$ -nm perpendicular mtj switching at sub-3-ns pulses below $100\ \mu$ a for high-performance embedded stt-mram for sub-20-nm cmos. *IEEE Transactions on Electron Devices* **64**, 427–431 (2017).

208. Torrezan, A. C., Strachan, J. P., Medeiros-Ribeiro, G. & Williams, R. S. Sub-nanosecond switching of a tantalum oxide memristor. *Nanotechnology* **22**, 485203, DOI: 10.1088/0957-4484/22/48/485203 (2011).

209. Jan, G. *et al.* Demonstration of ultra-low voltage and ultra low power stt-mram designed for compatibility with 0x node embedded llc applications. In *2018 IEEE Symposium on VLSI Technology*, 65–66 (2018).

210. Choi, B. J. *et al.* High-speed and low-energy nitride memristors. *Adv. Funct. Mater.* **26**, 5290–5296, DOI: 10.1002/adfm.201600680 (2016). https://onlinelibrary.wiley.com/doi/pdf/10.1002/adfm.201600680.

211. Kitagawa, E. *et al.* Impact of ultra low power and fast write operation of advanced perpendicular mtj on power reduction for high-performance mobile cpu. In *2012 International Electron Devices Meeting*, 29.4.1–29.4.4 (2012).

212. Francois, T. *et al.* Demonstration of beol-compatible ferroelectric hf0.5zr0.5o2 scaled feram co-integrated with 130nm cmos for embedded nvm applications. In *2019 IEEE International Electron Devices Meeting (IEDM)*, 15.7.1–15.7.4 (2019).

213. Luo, Q. *et al.* Super non-linear rram with ultra-low power for 3d vertical nano-crossbar arrays. *Nanoscale* **8**, 15629–15636, DOI: 10.1039/C6NR02029A (2016).

214. De Sandre, G. *et al.* A 90nm 4Mb embedded phase-change memory with 1.2V 12ns read access time and 1MB/s write throughput. In *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, 268–269 (2010).

215. Bruno, F. Y. *et al.* Millionfold resistance change in ferroelectric tunnel junctions based on nickelate electrodes. *Adv. Electron. Mater.* **2**, 1500245, DOI: 10.1002/aelm.201500245 (2016). https://onlinelibrary.wiley.com/doi/pdf/10.1002/aelm.201500245.

216. Kang, C.-F. *et al.* Self-formed conductive nanofilaments in (bi, mn)ox for ultralow-power memory devices. *Nano Energy* **13**, 283 – 290, DOI: https://doi.org/10.1016/j.nanoen.2015.02.033 (2015).

217. Xiong, F., Liao, A. D., Estrada, D. & Pop, E. Low-power switching of phase-change materials with carbon nanotube electrodes. *Science* **332**, 568–570, DOI: 10.1126/science.1201938 (2011). https://science.sciencemag.org/content/332/6029/568.full.pdf.

218. Wang, W. *et al.* Volatile Resistive Switching Memory Based on Ag Ion Drift/Diffusion—Part II: Compact Modeling. *IEEE Transactions on Electron Devices* **66**, 3802–3808, DOI: 10.1109/TED.2019.2928888 (2019).

219. Shi, Q., Wang, J., Aziz, I. & Lee, P. S. Stretchable and Wearable Resistive Switching Random-Access Memory. *Adv. Intell. Syst.* **2000007**, 2000007, DOI: 10.1002/aisy.202000007 (2020).

220. Shang, J. *et al.* Highly flexible resistive switching memory based on amorphous-nanocrystalline hafnium oxide films. *Nanoscale* **9**, 7037–7046, DOI: 10.1039/C6NR08687J (2017).

221. Dang, B. *et al.* Physically Transient Memristor Synapse Based on Embedding Magnesium Nanolayer in Oxide for Security Neuromorphic Electronics. *IEEE Electron Device Lett.* **40**, 1265–1268, DOI: 10.1109/LED.2019.2921322 (2019).

222. Mehonic, A. & Kenyon, A. J. Emulating the Electrical Activity of the Neuron Using a Silicon Oxide RRAM Cell. *Front. Neurosci.* **10**, 57, DOI: 10.3389/fnins.2016.00057 (2016).

223. Wang, W. *et al.* Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses. *Sci. Adv.* **4**, eaat4752, DOI: 10.1126/sciadv.aat4752 (2018).

224. Tuma, T., Pantazi, A., Le Gallo, M., Sebastian, A. & Eleftheriou, E. Stochastic phase-change neurons. *Nat. Nanotechnol.* **11**, 693–699, DOI: 10.1038/nnano.2016.70 (2016).

225. Suresh, B. *et al.* Simulation of integrate-and-fire neuron circuits using hfo2-based ferroelectric field effect transistors. In *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 229–232 (2019).

226. Kwon, M.-W. *et al.* Integrate-and-fire neuron circuit using positive feedback field effect transistor for low power operation. *J. Appl. Phys.* **124**, 152107 (2018).

227. Wang, Z. *et al.* Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **1**, 137–145 (2018).

228. Wang, Z., Ambrogio, S., Balatti, S. & Ielmini, D. A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning in neuromorphic systems. *Front. Neurosci.* **8**, 1–11, DOI: 10.3389/fnins.2014.00438 (2015).

229. Covi, E. *et al.* Analog Memristive Synapse in Spiking Networks Implementing Unsupervised Learning. *Frontiers in Neuroscience* **10**, 482, DOI: 10.3389/fnins.2016.00482 (2016).

230. Mulaosmanovic, H. *et al.* Novel ferroelectric fet based synapse for neuromorphic systems. In *2017 Symposium on VLSI Technology*, T176–T177 (IEEE, 2017).

231. Covi, E. *et al.* Spike-driven threshold-based learning with memristive synapses and neuromorphic silicon neurons. *J. Phys. D: Appl. Phys.* **51**, 344003 (2018).

232. Pedretti, G. *et al.* Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity. *Sci. Reports* **7**, 5288, DOI: 10.1038/s41598-017-05480-0 (2017).

233. Prezioso, M. *et al.* Spike-timing-dependent plasticity learning of coincidence detection with passively integrated memristive circuits. *Nat. Commun.* **9**, 5311, DOI: 10.1038/s41467-018-07757-y (2018).

234. Sebastian, A. *et al.* Temporal correlation detection using computational phase-change memory. *Nat. Commun.* **8**, 1115, DOI: 10.1038/s41467-017-01481-9 (2017). 1706.00511.

235. Wang, W. *et al.* Computing of temporal information in spiking neural networks with ReRAM synapses. *Faraday Discuss.* **213**, 453–469, DOI: 10.1039/C8FD00097B (2019).

236. Zhu, X., Du, C., Jeong, Y. & Lu, W. D. Emulation of synaptic metaplasticity in memristors. *Nanoscale* **9**, 45–51, DOI: 10.1039/C6NR08024C (2017).

237. Wu, Q. *et al.* Full imitation of synaptic metaplasticity based on memristor devices. *Nanoscale* **10**, 5875–5881, DOI: 10.1039/C8NR00222C (2018).

238. Burr, G. W. *et al.* Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element. *IEEE Transactions on Electron Devices* **62**, 3498–3507, DOI: 10.1109/TED.2015.2439635 (2015).

239. Yao, P. *et al.* Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646, DOI: 10.1038/s41586-020-1942-4 (2020).

240. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Natl. Acad. Sci. United States Am.* **79**, 2554–2558, DOI: 10.1073/pnas.79.8.2554 (1982).

241. Duan, S., Hu, X., Dong, Z., Wang, L. & Mazumder, P. Memristor-Based Cellular Nonlinear/Neural Network: Design, Analysis, and Applications. *IEEE Transactions on Neural Networks Learn. Syst.* **26**, 1202–1213, DOI: 10.1109/TNNLS.2014.2334701 (2015).

242. Romera, M. *et al.* Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).

243. Milo, V., Ielmini, D. & Chicca, E. Attractor networks and associative memories with STDP learning in RRAM synapses. In *2017 IEEE International Electron Devices Meeting (IEDM)*, 11.2.1–11.2.4 (2017).

244. Wang, Y., Yu, L., Wu, S., Huang, R. & Yang, Y. Memristor-Based Biologically Plausible Memory Based on Discrete and Continuous Attractor Networks for Neuromorphic Systems. *Adv. Intell. Syst.* **2**, 2000001, DOI: 10.1002/aisy.202000001 (2020).

245. 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**, DOI: 10.1126/sciadv.1700849 (2017). https://advances.sciencemag.org/content/3/10/e1700849.full.pdf.

246. Park, S. *et al.* Electronic system with memristive synapses for pattern recognition. *Sci. Reports* **5**, 1–9, DOI: 10.1038/srep10123 (2015).

247. Sheridan, P. M., Du, C. & Lu, W. D. Feature extraction using memristor networks. *IEEE Transactions on Neural Networks Learn. Syst.* **27**, 2327–2336 (2016).

248. Krestinskaya, O. & James, A. P. Feature extraction without learning in an analog spatial pooler memristive-cmos circuit design of hierarchical temporal memory. *Analog. Integr. Circuits Signal Process.* **95**, 457–465, DOI: 10.1007/s10470-018-1161-1 (2018).

249. Serb, A. *et al.* Memristive synapses connect brain and silicon spiking neurons. *Sci. Reports* **10**, 2590, DOI: 10.1038/s41598-020-58831-9 (2020).

250. Saleh, Q., Merkel, C., Kudithipudi, D. & Wysocki, B. Memristive computational architecture of an echo state network for real-time speech-emotion recognition. In *2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, 1–5 (2015).

251. Kudithipudi, D., Saleh, Q., Merkel, C., Thesing, J. & Wysocki, B. Design and analysis of a neuromemristive reservoir computing architecture for biosignal processing. *Front. Neurosci.* **9**, 502, DOI: 10.3389/fnins.2015.00502 (2016).

252. Zhu, S., Wang, L. & Duan, S. Memristive pulse coupled neural network with applications in medical image processing. *Neurocomputing* **227**, 149 – 157, DOI: https://doi.org/10.1016/j.neucom.2016.07.068 (2017).

253. Tzouvadaki, I., Tuoheti, A., De Micheli, G., Demarchi, D. & Carrara, S. Portable memristive biosensing system as effective point-of-care device for cancer diagnostics. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5 (2018).

254. Gokmen, T. & Vlasov, Y. Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices: Design Considerations. *Front. Neurosci.* **10**, 333, DOI: 10.3389/fnins.2016.00333 (2016).

255. Cai, F. *et al.* A fully integrated reprogrammable memristor–CMOS system for efficient multiply–accumulate operations. *Nat. Electron.* **2**, 290–299, DOI: 10.1038/s41928-019-0270-x (2019).

256. Woo, J. *et al.* Improved synaptic behavior under identical pulses using alox/hfo2 bilayer rram array for neuromorphic systems. *IEEE Electron Device Lett.* **37**, 994–997 (2016).

257. Wang, Z. *et al.* Engineering incremental resistive switching in taox based memristors for brain-inspired computing. *Nanoscale* **8**, 14015–14022, DOI: 10.1039/C6NR00476H (2016).

258. Mahmoodi, M. R., Prezioso, M. & Strukov, D. B. Versatile stochastic dot product circuits based on nonvolatile memories for high performance neurocomputing and neurooptimization. *Nat. Commun.* **10**, 5113, DOI: 10.1038/s41467-019-13103-7 (2019).

259. Cai, F. *et al.* Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks. *Nat. Electron.* **3**, 1–10, DOI: 10.1038/s41928-020-0436-6 (2020).

260. Midya, R. *et al.* Reservoir Computing Using Diffusive Memristors. *Adv. Intell. Syst.* **1**, 1900084, DOI: 10.1002/aisy.201900084 (2019).

261. Yang-Scharlotta, J., Fazio, M., Amrbar, M., White, M. & Sheldon, D. Reliability characterization of a commercial taox-based reram. In *IEEE IIRW*, 131–134 (2014).

262. Hayakawa, Y., Himeno, A., Yasuhara, R., Boullart, W. *et al.* Highly reliable taox reram with centralized filament for 28-nm embedded application. In *VLSI Technology*, T14–T15 (2015).

263. Hirtzlin, T. *et al.* Digital biologically plausible implementation of binarized neural networks with differential hafnium oxide resistive memory arrays. *Front. Neurosci.* **13**, 1383, DOI: 10.3389/fnins.2019.01383 (2020).

264. Chen, W.-H. *et al.* Cmos-integrated memristive non-volatile computing-in-memory for ai edge processors. *Nat. Electron.* **2**, 420–428 (2019).

265. Valentian, A. *et al.* Fully integrated spiking neural network with analog neurons and rram synapses. In *2019 IEEE International Electron Devices Meeting (IEDM)*, 14.13.1–14.13.4 (2019).

266. Shulaker, M. M. *et al.* Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **547**, 74–78 (2017).

267. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Networks* **113**, 54–71 (2019).

268. Brivio, S. *et al.* Extended memory lifetime in spiking neural networks employing memristive synapses with nonlinear conductance dynamics. *Nanotechnology* **30**, 015102 (2018).

269. Frascaroli, J., Brivio, S., Covi, E. & Spiga, S. Evidence of soft bound behaviour in analogue memristive devices for neuromorphic computing. *Sci. Reports* **8**, 7178, DOI: 10.1038/s41598-018-25376-x (2018).

270. Muñoz-Martín, I. *et al.* Unsupervised learning to overcome catastrophic forgetting in neural networks. *IEEE J. on Explor. Solid-State Comput. Devices Circuits* **5**, 58–66 (2019).