Integration of nanoscale memristor synapses in neuromorphic computing architectures

Giacomo Indiveri[a,1]**, Bernabe Linares-Barranco**[b]**, Robert Legenstein**[c]**, George Deligeorgis**[d]**, and Themistoklis Prodromakis**[e]

[a] Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland
[b]Instituto Microelectronica Sevilla (IMSE-CNM-CSIC), Sevilla, Spain
[c]Institute for Theoretical Computer Science, Graz University of Technology, Graz, Austria
[d]CNRS-LAAS and Univ de Toulouse, 7 avenue du colonel Roche, F-31400 Toulouse, France
[e]Center for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom

E-mail: [1]`giacomo@ethz.ch`

Abstract. Conventional neuro-computing architectures and artificial neural networks have often been developed with no or loose connections to neuroscience. As a consequence, they have largely ignored key features of biological neural processing systems, such as their extremely low-power consumption features or their ability to carry out robust and efficient computation using massively parallel arrays of limited precision, highly variable, and unreliable components. Recent developments in nano-technologies are making available extremely compact and low-power, but also variable and unreliable solid-state devices that can potentially extend the offerings of availing CMOS technologies. In particular, memristors are regarded as a promising solution for modeling key features of biological synapses due to their nanoscale dimensions, their capacity to store multiple bits of information per element and the low energy required to write distinct states. In this paper, we first review the neuro- and neuromorphic-computing approaches that can best exploit the properties of memristor and-scale devices, and then propose a novel hybrid memristor-CMOS neuromorphic circuit which represents a radical departure from conventional neuro-computing approaches, as it uses memristors to directly emulate the biophysics and temporal dynamics of real synapses. We point out the differences between the use of memristors in conventional neuro-computing architectures and the hybrid memristor-CMOS circuit proposed, and argue how this circuit represents an ideal building block for implementing brain-inspired probabilistic computing paradigms that are robust to variability and fault-tolerant by design.

1. Introduction

The idea of linking the type of information processing that takes place in the brain with theories of computation and computer science (something commonly referred to as *neuro-computing*) dates back to the origins of computer science itself [1, 2]. Neuro-computing has been very popular in the past [3, 4], eventually leading to the development of abstract artificial neural networks implemented on digital computers, useful for solving a wide variety of practical problems [5, 6, 7, 8, 9]. However, the field of *neuromorphic engineering* is a much younger one [10]. This field has been mainly concerned with hardware implementations of neural processing and sensory-motor systems built using Very Large Scale Integration (VLSI) electronic circuits that exploit the physics of silicon to reproduce directly the biophysical processes that underlie neural computation in real neural systems. Originally, the term "neuromorphic" (coined by Carver Mead in 1990 [11]) was used to describe systems comprising analog integrated circuits, fabricated using standard Complementary Metal Oxide Semiconductor (CMOS) processes. In recent times, however, the use of this term has been extended to refer to hybrid analog/digital electronic systems, built using different types of technologies.

Indeed, both artificial neural networks and neuromorphic computing architectures are now receiving renewed attention thanks to progress in Information and Communication Technologys (ICTs) and to the advent of new promising nanotechnologies. Some of present day neuro-computing approaches attempt to model the fine details of neural computation using standard technologies. For example, the *Blue Brain* project, launched in 2005, makes use of a 126kW Blue Gene/P IBM supercomputer to run software that simulates with great biological accuracy the operations of neurons and synapses of a rat neocortical column [12]. Similarly, the *BrainScaleS* EU-FET FP7 project aims to develop a custom neural supercomputer by integrating standard CMOS analog and digital VLSI circuits on full silicon wafers to implement about 262 thousand Integrate-and-Fire (I&F) neurons and 67 million synapses [13]. Although configurable, the neuron and synapse models are hardwired in the silicon wafers, and the hardware operates about 10000 times faster than real biology, with each wafer consuming about 1kW power, excluding all external components. Another large-scale neuro-computing project based on conventional technology is the *SpiNNaker* project [14]. The SpiN-Naker is a distributed computer, which interconnects conventional multiple integer precision multi ARM core chips via a custom communication framework. Each SpiNNaker package contains a chip with 18 ARM9 Central Processing Units (CPUs) on it, and a memory chip of 128 Mbyte Synchronous Dynamic Random Access Memory (DRAM). Each CPU can simulate different neuron and synapse models. If endowed with simple synapse models, a single SpiNNaker device ARM core can simulate the activity of about 1000 neuron in real time. More complex synapse models (e.g. with learning mechanisms) would use up more resources and decrease the number of neurons that could be simulated in real-time. The latest SpiNNaker board contains 47 of these packages, and the aim is to assemble 1200 of these boards. A full SpiNNaker system of this size

would consume about $90\,kW$. The implementation of custom large-scale spiking neural network hardware simulation engines is being investigated also by industrial research groups. For example, the IBM group led by D.S. Modha recently proposed a digital "neurosynaptic core" chip integrated using a standard $45\,nm$ Silicon on Insulator (SOI) process [15]. The chip comprises 256 digital I&F neurons, with 1024×256 binary valued synapses, configured via a Static Random Access Memory (SRAM) cross-bar array, and uses an asynchronous event-driven design to route spikes from neurons to synapses. The goal is to eventually integrate many of these cores onto a single chip, and to assemble many multi-core chips together, to simulate networks of simplified spiking neurons with human-brain dimensions (i.e. approximately 10^{10} neurons and 10^{14} synapses) in real-time. In the mean time, IBM simulated 2.084 billion neurosynaptic cores containing 53×10^{10} neurons and 1.37×10^{14} synapses in software on the Lawrence Livermore National Lab Sequoia supercomputer (96 Blue Gene/Q racks), running $1542\times$ slower than real time [16], and dissipating $7.9\,MW$. A diametrically opposite approach is represented by the *Neurogrid* system [17]. This system comprises an array of sixteen $12 \times 14\,mm^2$ chips, each integrating mixed analog neuromorphic neuron and synapse circuits with digital asynchronous event routing logic. The chips are assembled on a $16.5 \times 19\,cm^2$ Printed Circuit Board (PCB), and the whole system can model over one million neurons connected by billions of synapses in real-time, and using only about $3\,W$ of power [18]. As opposed to the neuro-computing approaches that are mainly concerned with fast and large simulations of spiking neural networks, the Neurogrid has been designed following the original neuromorphic approach, exploiting the characteristics of CMOS VLSI technology to directly emulate the biophysics and the connectivity of cortical circuits. In particular, the Neurogrid network topology is structured by the data and results obtained from neuro-anatomical studies of the mammalian cortex. While offering less flexibility in terms of connectivity patterns and types of synapse/neuron models that can be implemented, the Neurogrid is much more compact and dissipates orders of magnitude less power than the other neuro-computing approaches described above. All these approaches have in common the goal of attempting to simulate large numbers of neurons, or as in the case of Neurogrid, to physically emulate them with fine detail.

Irrespective of the approach followed, nanoscale synapse technologies and devices have the potential to greatly improve circuit integration densities and to substantially reduce power-dissipation in these systems. Indeed, recent trends in nanoelectronics have been investigating emerging low-power nanoscale devices for extending standard CMOS technologies beyond the current state-of-art [19]. In particular, Resistive Random Access Memory (ReRAM) is regarded as a promising technology for establishing next-generation non-volatile memory cells [20], due to their infinitesimal dimensions, their capacity to store multiple bits of information per element and the minuscule energy required to write distinct states. The factors driving this growth are attributed to the devices' simple (two terminals) and infinitesimal structure (state-of-art is down to $10\times10\,nm^2$ [21]) and ultra-low power consumption ($< 50\,fJ/bit$) that so far are unmatched by conventional VLSI circuits.

Various proposals have already been made for leveraging basic nanoscale ReRAM attributes in reconfigurable architectures [22], neuro-computing [23] and even artificial synapses [24, 25, 26, 27, 28]. However the greatest potential of these nanoscale devices lies in the wide range of interesting physical properties they possess. Neuromorphic systems can harness the interesting physics being discovered in these new nanodevices to emulate the biophysics of real synapses and neurons and reproduce relevant computational primitives, such as state-dependent conductance changes, multi-level stability and stochastic state changes in large-scale artificial neural systems.

In this paper we first describe how nanoscale synaptic devices can be integrated into neuro-computing architectures to build large-scale neural networks, and then propose a new hybrid memristor-CMOS neuromorphic circuit that emulates the behavior of real synapses, including their temporal dynamics aspects, for exploring and understanding the principles of neural computation and eventually building brain-inspired computing systems.

2. Solid-state memristors

ReRAM cells are nowadays classified as being memory-resistors [29], or memristors for short, that have first been conceptually conceived in 1971 by Leon Chua [30]; with the first biomimetic applications presented at the same time. The functional signature of memristors is a pinched hysteresis loop in the current-voltage (i-v) domain when excited by a bipolar periodic stimulus [31]. Such hysteresis is typically noticed for all kind of devices/materials in support of a discharge phenomenon that possess certain inertia, causing the value of a physical property to lag behind changes in the mechanism causing it, and has been common both to large scale [32] as well as nanoscale dissipative devices [33].

2.1. Emerging nanodevices as synapse mimetics

The analogy of memristors and chemical synapses is made on the basis that synaptic dynamics depend upon the discharge of neurotransmitters within a synaptic cleft (see Fig. 1a), in a similar fashion that "ionic species" can be displaced within any inorganic barrier (see Fig. 1b). TiO_2-based memristor models [33, 34] hypothesized that solid-state devices comprise a mixture of TiO_2 phases, a stoichiometric and a reduced one ($TiO_2 - x$), that can facilitate distinct resistive states via controlling the displacement of oxygen vacancies and thus the extent of the two phases. More recently however it was demonstrated that substantial resistive switching is only viable through the formation and annihilation of continuous conductive percolation channels [35] that extend across the whole active region of a device, shorting the top and bottom electrodes; no matter what the underlying physical mechanism is.

An example of I-V characteristics of TiO_2-based memristors is depicted in Fig. 2a. In this example, consecutive positive voltage sweeps cause any of the cross-bar type
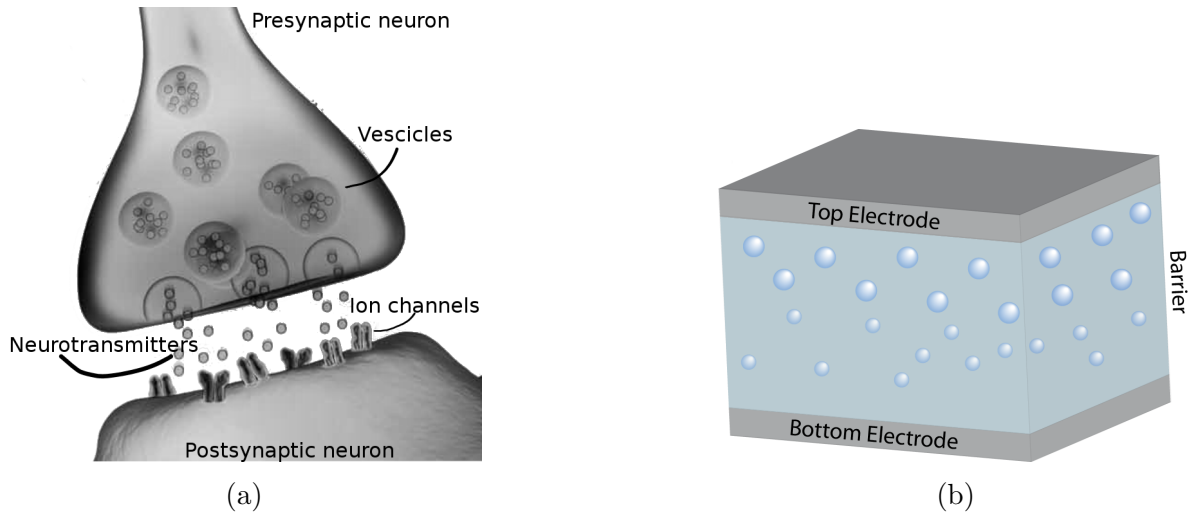
Figure 1: (a) Cross-section of a chemical synapse, illustrating the discharge of neurotransmitters within a synaptic cleft originating from a pre-synaptic neuron. (b) Schematic representation of solid-state memristors where ionic species can be displaced within a device's insulating medium, transcribing distinct resistive states, by application of electrical stimuli on the top or bottom electrodes of the device.

devices [36] shown in the inset of Fig. 2a to switch from a High-Resistive State (HRS) to Low-Resistive States (LRSs). When the polarity of the voltage sweeps is however inverted, the opposite trend occurs, i.e. the device toggles from LRS to HRS consecutively (as indicated by the corresponding arrows). These measured results are consistent with analogous ones proposed by other research groups [37, 38, 39] and demonstrate the devices' capacity for storing a multitude of resistive states per unit cell, with the programming depending on the biasing history. This is further demonstrated in Fig. 2b, by applying individual pulses of -3 V in amplitude and $1\,\mu$sec long for programming a single memristor at distinct non-volatile resistive states. In this scenario, the solid-state memristor emulates the behavior of a depressing synapse [40, 41]; the inverse, i.e. short-term potentiation is also achievable by alternating the polarity of the employed pulsing scheme.

The development of nanoscale dynamic computation elements may notably benefit the establishment of neuromorphic architectures. This technology adds substantially on computation functionality, due to the rate-dependency of the underlying physical switching mechanisms. At the same time it can facilitate unprecedented complexity due to the capacity of storing and processing spiking events locally. Moreover, exploiting the nanoscale dimensions and architecture simplicity of solid-state memristor implementations could substantially augment the number of cells per unit area, effectively enhancing the systems' redundancy for tolerating issues that could stem from device mismatch and low-yields [42].
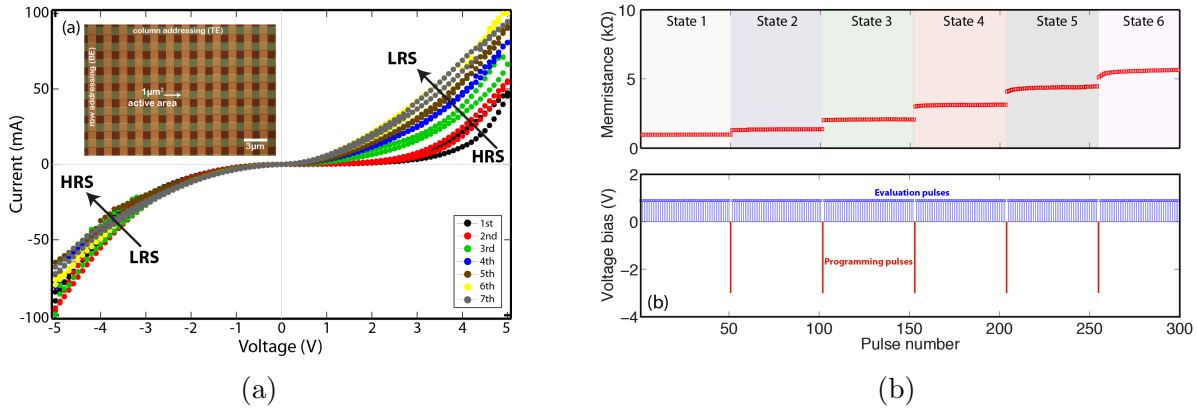
(a)

(b)

Figure 2: Characterization of a TiO_2-based solid-state memristor. (a) I-V characteristics for consecutive voltage sweeping. Positive (negative) biasing renders an increase (decrease) in the device's conductance. Inset of (a) depicts a 25×25 array crossbar type memristors comprising of TiO_2 active areas of $1 \times 1\mu m^2$. These cells can be programmed at distinct resistive states as shown in (b) by employing -3V and $1\mu sec$ wide pulses, while evaluation of the device's states is performed at 0.9V.

2.2. Memristor scaling

Resistive memory scaling has been intensively investigated for realization of nanosized ReRAM [43]. In principle memristors may be scaled aggressively well below conventional RAM cells due to their simplicity: fabrication-wise memristors typically rely on a Metal Insulator Metal (MIM) structure. The memristor action occurs in the insulating material. Scaling down the thickness of such a material will reduce both the required set voltage as well as the read voltage used during operation. In this context, thickness figures of a few nano meters have been demonstrated and operating voltages below 1 V have been shown [44] with a switching energy of a few fJ [45]. Furthermore, reducing the active device area by down-scaling the electrodes leads to current scaling, as well as increased device density. Both of these effects are favorable for high complexity circuits.

Currently even though single memristor devices as small as 10×10 nm have been demonstrated [21], cross-bar arrays are the most commonly used architecture [46, 36] to organize large numbers of individually addressable memristive synapses in a reduced space. In Fig. 3 we show a large array of nanoscale memristors that we fabricated using electron beam lithography. This array consists of a continuous Pt bottom electrode and an active layer deposited by Sputtering. Subsequently, several arrays of nano-memristors with a size ranging from 20 to 50 nm were defined using E-beam lithography on PMMA and lift-off of the top Platinum electrode. The array shown here comprises 256×256 devices with a periodicity of 200 nm. To access each individual device a conductive Atomic Force Microscope (AFM) tip was used. Such a structure has been used to study the variability of the fabricated devices. Using E-beam lithography for both the top and bottom electrodes a fully interconnected cross bar structure with similar size and pitch may be fabricated.
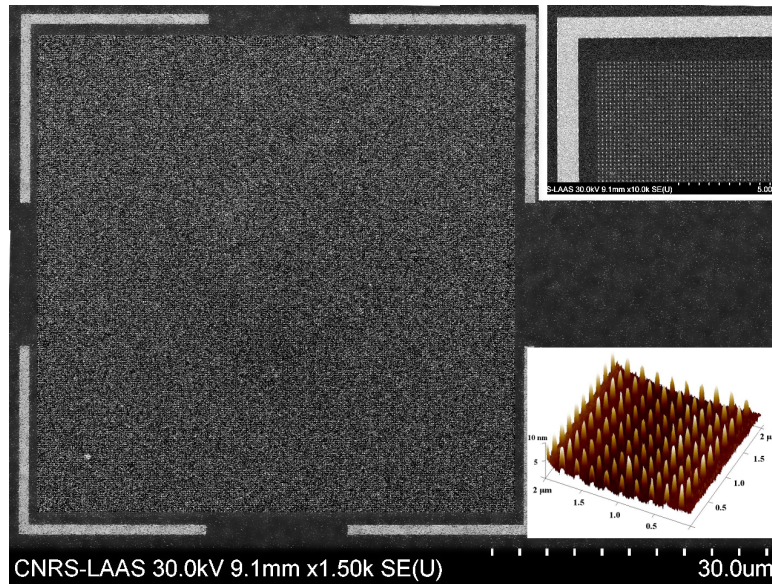
Figure 3: SEM micrograph of a large nanosized memristor array. Top inset shows a zoom-in of the top left corner where the individual devices are distinguished. Bottom left inset shows an AFM image of a small part of the array. Individual devices are addressed by placing a conductive AFM tip on the top electrode.

3. Memristor-based neuro-computing architectures

Memristive devices have been proposed as analogs of biological synapses. Indeed, memristors could implement very compact but abstract models of synapses, for example representing a binary "potentiated" or "depressed" state, or storing an analog "synaptic weight" value [47]. In this framework, they could be integrated in large and dense cross-bar arrays [48] to connect large numbers of silicon neurons [49], and used in a way to implement spike-based learning mechanisms that change their local conductance.

In [25, 24] the authors proposed a scheme where neurons can drive memristive synapses to implement a Spike-Timing Dependent Plasticity (STDP) [50] learning scheme by generating single pairs of pre- and post-synaptic spikes in a fully asynchronous manner, without any need for global or local synchronization, thus solving some of the problems that existed with previously proposed learning schemes [51, 28]. The main idea is the following: when no spike is generated, each neuron maintains a constant reference voltage at both its input and output terminals. During spike generation, each neuron forces a pre-shaped voltage waveform at both its input and output terminals, as shown in Fig. 4a, to update the synaptic weight value stored in the memristor state. Since memristors change their resistance when the voltages at their terminals exceed some defined thresholds, it is possible to obtain arbitrary STDP weight update functions, including biologically plausible ones, as the one shown in Fig. 4b [50]. Moreover by properly shaping the spike wave-forms of both pre- and post-synaptic spikes it is possible to change the form of the STDP learning function, or to even make it evolve in time as
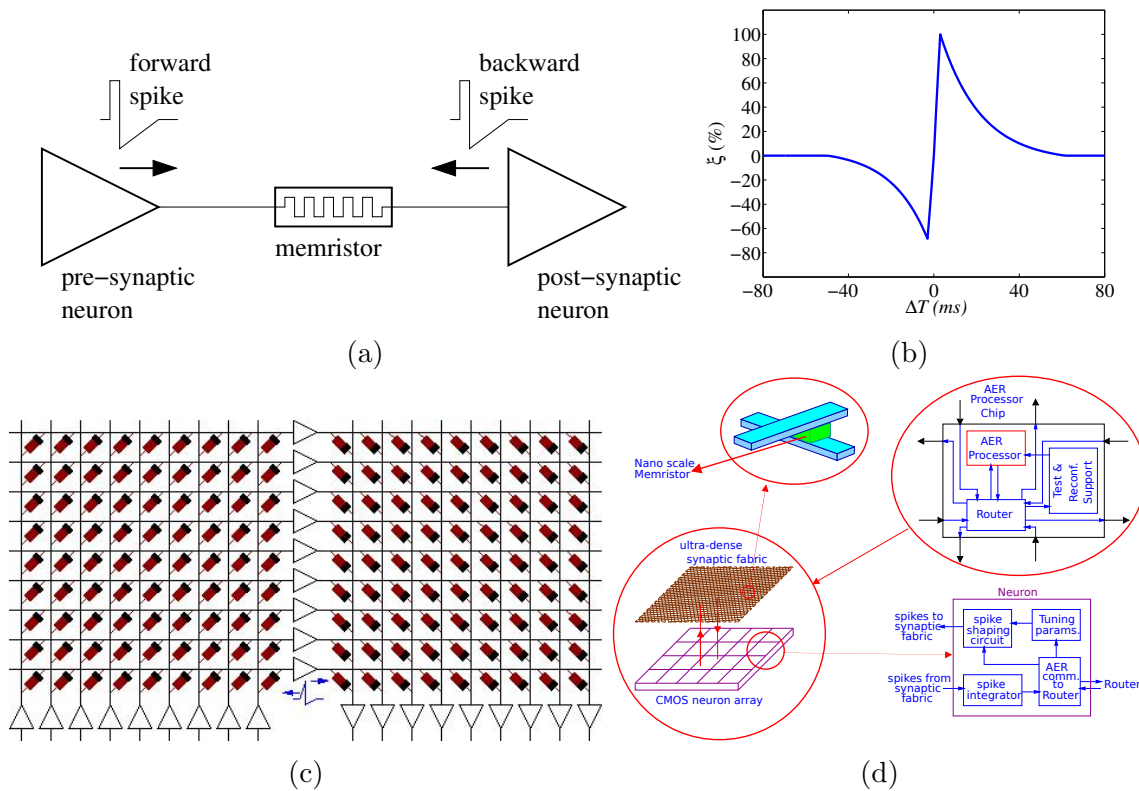
Figure 4: Single memristor synapse concept.(a) One Memristor synapse with pre- and post-synaptic pulse-shaping neuron circuits. (b) Example of a STDP weight update learning function $\xi(\Delta T)$, where ΔT represents the difference between the timing of the post-synaptic and pre-synaptic spikes. (c) Circuit architecture comprising three neuron layers connected by means of synaptic crossbars. (d) Hybrid memristor/CMOS neurons and AER 2D chip architecture for spike/event routing and processing. Parts of this figure were adapted from [24].

learning progresses [52, 25]. Fully interconnected or partially interconnected synaptic crossbar arrays, as illustrated in Fig. 4c, could facilitate hierarchical learning neural network architectures. Since there is no need for global synchronization, this approach could be extended to multi-chip architectures that transmit spikes across chip boundaries using fully asynchronous timing. For example, a common asynchronous communication protocol that has been used in neuromorphic systems is based on the Address Event Representation (AER) [53, 54]. In this representation, each spiking neuron is assigned an address, and when the neuron fires an address-event is put on a digital bus, at the time that the spike is emitted. In this way time represent itself, and information is encoded in real-time, in the inter-spike intervals. By further exploiting hybrid CMOS/memristor chip fabrication techniques [55], this approach could be easily scaled up to arbitrarily large networks (e.g., see Fig. 4d). Following this approach each neuron processor would be placed in a 2D grid fully, or partially interconnected through memristors. Each neuron would perform incoming spike aggregation, provide the desired pre- and post-

synaptic (programmable) spike waveforms, and communicate incoming and outgoing spikes through AER communication circuitry. Using state-of-the-art CMOS technology, it is quite realistic to provide in the order of a million such neurons per chip with about 10^4 synapses per neuron. For example, by using present day 40 nm CMOS technology it is quite realistic to fit a neuron within a $10\mu m \times 10\mu m$ area. This way, a chip of about $1cm^2$ could host of the order of one million neurons. At the same time, for the nano wire fabric deposited on top of CMOS structures, present day technology can easily provide nano wires of 100nm pitch [21]. This would allow to integrate about 10^4 synapses on top of the area occupied by each CMOS neuron. Similarly, at the PCB level, it is possible to envisage that a 100-chip PCB could host about 10^8 neurons, and 40 of these PCBs would emulate 4 billion neurons. In these large-scale systems the bottleneck is largely given by the spike or event communication limits. To cope with these limits such chips would inter-communicate through nearest neighbors, exploiting 2D-grid network-on-chip (NoC) and network-on-board (NoB) principles. For example, in [56] the authors proposed a very efficient multi-chip inter-communication scheme that distributes event traffic over a 2D mesh network locally within each board through inter-chip high speed buses. Reconfigurability and flexibility would be ensured by defining the system architecture and topology through in-chip routing tables. Additionally, by arranging the neurons within each chip in a local 2D mesh with in-chip inter-layer event communication, it is possible to keep most of the event traffic inside the chips. At the board level, the 2D mesh scheme would allow for a total inter-chip traffic in the order of $E_v = 4N_{ch} \times E_{pp}$, where $N_{ch} = 100$ is the number of chips per board, E_{pp} is the maximum event bandwidth per inter-chip bus (which we may assume to be around 100 Meps - mega events per second), and 4 reflects the fact that each chip is connected to its four nearest neighbors [56]. With these numbers, the maximum traffic per board would be in the order of $E_v \approx 4 \times 10^{10} eps$, which is about 400 eps per neuron just for inter-chip event exchange. In practice, inter-board traffic could be much sparser, if the system is partitioned efficiently. Such numbers are quite realistic for present day CMOS technology, and the approach is scalable. Regarding power consumption of the communication overhead, we can use as reference some recent developments for event-based fully bit-serial inter-chip transmission schemes over differential micro strips [57, 56], where consumption is proportional to communication event rate. Each link would consume in the order of 40 mA at 10 Meps rate (this includes driver and receiver pad circuits [57] as well as serializers and deserializers [58]). If each neuron fires at an average rate of 1 Hz, and if each chip has 1 million neurons, the current consumption of the communication overhead would be about 4 mA per chip. If voltage supply is in the 1-2 V range, this translates into 4-8 mW per chip. For a 100 chip PCB the inter-chip communication overhead power consumption would thus be about 400-800 mW, for 1 Hz average neuron firing rate.

4. Neuromorphic and hybrid memristor-CMOS synapse circuits

We've shown how memristive devices and nano-technologies can be exploited to dramatically increase integration density and implement large-scale abstract neural networks. However to faithfully reproduce the function of real synapses, including their temporal dynamic properties, passive memristive devices would need to be interfaced to biophysically realistic CMOS circuits that follow the neuromorphic approach, as described in [10, 11]. On one hand, building physical implementations of circuits and materials that directly emulate the biophysics of real synapses and reproduce their detailed real-time dynamics is important for basic research in neuroscience, on the other, this neuromorphic approach can pave the way for creating an alternative non-von Neumann computing technology, based on massively parallel arrays of slow, unreliable, and highly variable, but also compact and extremely low-power solid-state components for building neuromorphic systems that can process sensory signals and interact with the user and the environment in real-time, and possibly carry out computation using the same principles used by the brain. Within this context, of massively parallel artificial neural processing elements, memory and computation are co-localized. Typically the amount of memory available per each "computing node" (synapse in our case) is limited and it is not possible to transfer and store partial results of a computation in large memory banks outside the processing array. Therefore, in order to efficiently process real-world biologically relevant sensory signals these types of neuromorphic systems must use circuits that have biologically plausible time constants (i.e., of the order of tens of milliseconds). In this way, in addition to being well matched to the signals they process, these systems will also be inherently synchronized with the real-world events they process and will be able to interact with the environment they operate in. But these types of time constants require very large capacitance and resistance values. For example, in order to obtain an equivalent RC time constant of 10 ms with a resistor even as large as $10\,\mathrm{M\Omega}$, it would be necessary to use a capacitor of 100 pF. In standard CMOS VLSI technology a synapse circuit with this RC element would require a prohibitively large area, and the advantages of large-scale integration would vanish. One elegant solution to this problem is to use current-mode design techniques [59] and log-domain *subthreshold* circuits [60, 61]. When Meal Oxide Semiconductor Field Effect Transistors (MOSFETs) are operated in the subthreshold domain, the main mechanism of carrier transport is that of diffusion [60], the same physical process that governs the flow of ions through proteic channels across neuron membranes. As a consequence, MOSFETs have an exponential relationship between gate-to-source voltage and drain current, and produce currents that range from femto- to nano-Ampères. In this domain it is possible to implement active VLSI analog filter circuits that have biologically realistic time-constants and that employ relatively small capacitors.

(a)   (b)
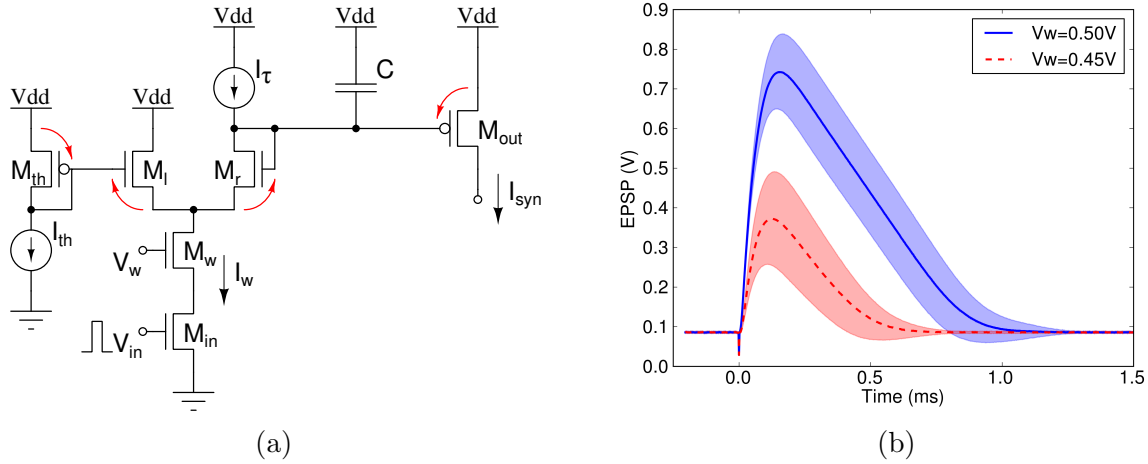
Figure 5: Neuromorphic electronic synapses (a) Log-domain DPI circuit diagram of an excitatory silicon synapse. Red arrows show the translinear loop considered to derive the circuit response. Input voltage spikes V_{in} are integrated by the circuit to produce post-synaptic currents I_{syn} with biologically faithful dynamics. (b) Experimental data showing the EPSP response of the circuit for two different settings of synaptic weight bias voltage V_w. The data was measured from the DPI synapses of 124 neurons, integrated on the same chip, with shared common bias settings. The dashed and solid lines represent the average response, while the shaded areas (standard deviation) indicate the extend of the device mismatch effect.

4.1. A CMOS neuromorphic synapse

An example of a compact circuit that can produce both linear dynamics with biologically plausible time constants as well as non-linear short-term plasticity effects analogous to those observed in real neurons and synapses is the Differential Pair Integrator (DPI) circuit [62] shown in Fig. 5a. It can be shown [63] that by exploiting the *translinear-principle* [64] across the loop of gate-to-source voltages highlighted in the figure, the circuit produces an output current I_{syn} with impulse response of the form:

$$\tau \frac{d}{dt} I_{syn} + I_{syn} = \frac{I_w I_{th}}{I_\tau}, \tag{1}$$

where $\tau \triangleq C U_T / \kappa I_\tau$ is the circuit time constant, κ the subthreshold slope factor [60], and $U_T = KT/q$ represents the thermal voltage. The currents I_w and I_{th} represent local synaptic weight and a global synaptic scaling gain terms, useful for implementing spike-based and homeostatic plasticity mechanisms [65, 66]. Therefore, by setting for example, $I_\tau = 5\,pA$, and assuming that $U_T = 25\,\text{mV}$ at room temperature, the capacitance required to implement a time constant of $10\,\text{ms}$ would be approximately $C = 1\,\text{pF}$. This can be implemented in a compact layout and allows the integration of large numbers of silicon synapses with realistic dynamics on a small VLSI chip. The same circuit of Fig. 5a can be used to implement elaborate models of spiking neurons, such as the

"Adaptive Exponential" (AdExp) I&F model [67, 49]. Small (minimum-size, of about $10\,\mu m^2$) prototype VLSI chips comprising of the order of thousands of neurons and synapses based on the DPI circuit have been already fabricated using a conservative $350\,nm$ CMOS technology [68]. The data of Fig. 5b shows the average response of a DPI synapse circuits measured from one of such chips [68]. The data represents the average Excitatory Post Synaptic Potential (EPSP) produced by 124 neurons in response to a single spike sent to the DPI synapses of each neuron. The shaded areas, representing the standard deviation, highlight the extent of variability present in these types of networks, due to device mismatch. The main role of the DPI circuit of Fig. 5a is to implement synaptic dynamics. Short-term plasticity, STDP learning, and homeostatic adaptation mechanisms can be, and have been, implemented by interfacing additional CMOS circuits to control the DPI V_w bias voltage, or to the I_{th} bias current [62, 69, 70]. Long-term storage of the V_w weights however requires additional power-consuming and area-expensive circuit solutions, such as floating gate circuits, or local Analog to Digital Converter (ADC) and SRAM cells.

4.2. A new hybrid memristor-CMOS neuromorphic synapse

Nano-electronic technologies offer a promising alternative solution for compact and low-power long-term storage of synaptic weights. The hybrid memristor-CMOS neuromorphic synapse circuit we propose here, shown in Fig. 6a, exploits these features to obtain at the same time dense integration of low-power long-term synaptic weight storage elements, and to emulate detailed synaptic biophysics for implementing relevant computational properties of neural systems.

The circuit depicted in Fig. 6a represents a possible implementation of a dense array of N synapses with independent weights but with the same, shared, temporal dynamics. Depending on their size, each memristor in Fig. 6a could represent a full synaptic contact, or an individual ion channel in the synaptic cleft (see also Fig. 1a). If the currently accepted model of filament formation in memristive devices is true, then down-scaled memristors should approach single filament bistable operation. While this is a severe limitation for classical neural network applications in which memristors are required to store analog synaptic weight values with some precision, it would actually provide a very compact physical medium for emulating the stochastic nature of the opening and closing of ion channels in biological synapses.

The shared temporal dynamics are implemented by the DPI circuit in the top part of Fig. 6a. Indeed, if this circuit is operated in its linear regime, it is possible to time-multiplex the contributions from all spiking inputs, thus requiring one single integrating element and saving precious silicon real-estate. The V_w bias voltage of this circuit is a global parameter that sets the maximum possible current that can be produced by each memristor upon the arrival of an input spike, while the memristor conductance modulates the current being produced by the synapse very much like conductance changes in real synapses affect the Excitatory Post Synaptic Currents (EPSCs) they
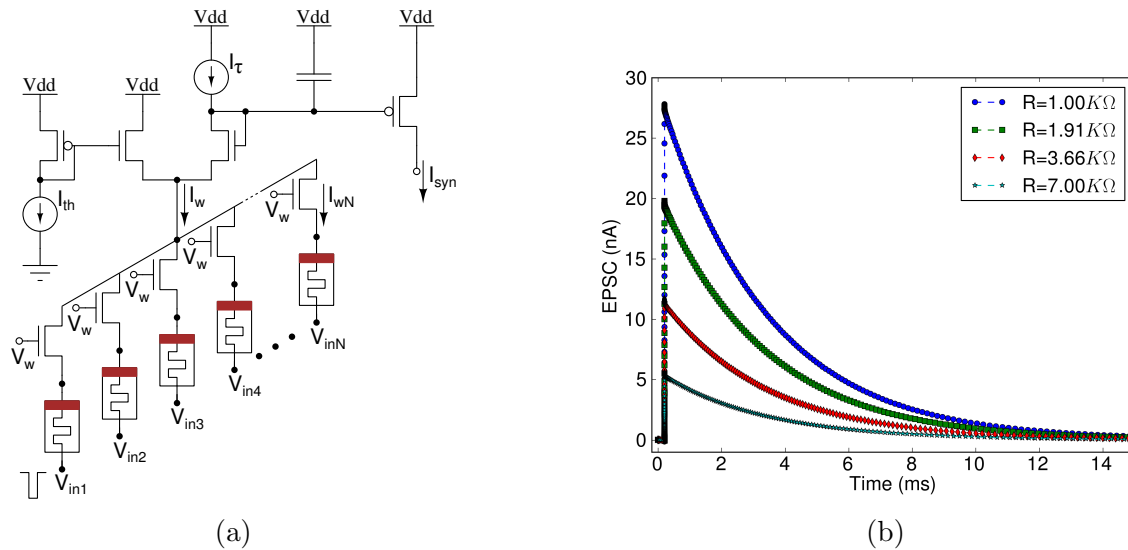
Figure 6: Neuromorphic memristive synapse. (a) Schematic circuit implementing an array of memristive synapses, with independent inputs and synaptic weights, but with shared temporal dynamics. (b) SPICE simulations of the circuit in Fig. 6a showing the output I_{syn} EPSC in response to a pre-synaptic input spike, for 4 different memristor conductance values.

produce. Larger memristor conductances, which represent a larger number of open proteic channels in real synapses, correspond to larger synaptic weights.

Figure 6b shows the results of SPICE simulations of the circuit in Fig. 6a, for a 180 nm CMOS process. The I_{thr} and I_τ current sources were implemented with p-type MOSFETs, biased to produce 2 pA and 10 pA respectively, and the V_w voltage bias was set to 700 mV. The data was obtained by simulating the response of one input memristive branch to a single input spike, while sweeping the memristor impedance from 1 KΩ to 7 KΩ. In these simulations we set the memristor in its LRS, and assumed we could modulate the value of the resistance to obtain four distinct analog states analogous to the ones measured experimentally in Fig. 2b. Of course the circuit supports also the operation of the memristor as a binary device, working in either the HRS state or the LRS one. This bi-stable mode of using the memristor would encode only an "on" or "off" synaptic state, but it would be more reliable and it is compatible with biologically plausible learning mechanisms, such as those proposed in [71], and implemented in [69]. The circuit of Fig. 6a shows only the circuit elements required for a "read" operation, i.e., an operation that stimulates the synapse to generate an EPSC with an amplitude set by the conductance of the memristor. Additional circuit elements would be required to change the value of the memristor's conductance, e.g., via learning protocols. However the complex circuitry controlling the learning mechanisms would be implemented at the Input/Output (I/O) periphery of the synaptic array, for example with pulse-shaping circuits and architectures analogous to the ones described in Section 3, or with circuits that check the state of the neuron and of it's recent spiking history, such as those

proposed in [61], and only a few additional compact elements would be required in each synapse to implement the weight update mechanisms.

5. Brain-inspired probabilistic computation

While memristors offer a compact and attractive solution for long-term storage of synaptic state, as done for example in Fig. 6, they are affected by a high degree of variability (e.g., much higher than the one measured for CMOS synapses in Fig. 5b). In addition, as memristors are scaled down, unreliable and stochastic behavior becomes unavoidable. The variability, stochasticity, and in general reliability issues that are starting to represent serious limiting factors for advanced computing technologies, do not seem to affect biological computing systems. Indeed, the brain is a highly stochastic system that operates using noisy and unreliable nanoscale elements. Rather than attempting to minimize the effect of variability in nanotechnologies, one alternative strategy, compatible with the neuromorphic approach, is to embrace variability and stochasticity and exploit these "features" to carry out robust brain-inspired probabilistic computation.

The fact that the brain can efficiently cope with a high degree of variability is evident at many levels: at the macroscopic level trial-to-trial variability is present for example in the arm trajectories of reaching movement tasks. It is interesting to note that the variability of the end position of the reaching movement is reduced, if the task requires to hit or touch a target with high accuracy [72]. Variability is evident at the level of cortical neurons: there is significant trial to trial variability in their responses to identical stimuli; it is evident also at the level of chemical synapses, where there is a high degree of stochasticity in the transmission of neurotransmitter molecules [73], from the pre-synaptic terminal to the post-synaptic one. The release probability of cortical synapses ranges from values of less than 1% to 100% [74]. This indicates that stochastic synaptic release may not merely be an unpleasant constraint of the molecular machinery but may rather be an important computational feature of cortical synapses.

What could be the computational benefit of using hardware affected by variability and stochasticity in biological and artificial computing systems? Recent advances in cognitive science demonstrated that human behavior can be described much better in the framework of *probabilistic inference* rather than in the framework of traditional "hard" logic inference [75], and encouraged the view that neuronal networks might directly implement a process of probabilistic inference [76]. In parallel, to this paradigm shift, research in machine learning has revealed that probabilistic inference is often much more appropriate for solving real-world problems, then hard logic [77]. The reason for this is that reasoning can seldom be based on full and exact knowledge in real-world situations. For example, the sensory data that a robot receives is often noisy and incomplete such that the current state of the environment can only partially be described. Probabilistic reasoning is a powerful tool to deal with such uncertain situations. Of course, exact probabilistic inference is still computationally intractable in general, but a number of

approximation schemes have been developed that work well in practice.

In probabilistic inference, the idea is to infer a set of unobserved variables (e.g., motor outputs, classification results, etc.) given a set of observed variables (evidence, e.g., sensory inputs), using known or learned probabilistic relationships among them. Specifically, if the distribution $P(\bar{x})$ describes the probabilistic relationships between the random variables $x_1, \ldots, x_n$, and if $x_1, \ldots, x_k$ of this distribution are observed, then one can infer a set of variables of interests $x_{k+1}, \ldots, x_{k+l}$ by determining the posterior probability $P(x_{k+1}, \ldots, x_{k+l} | x_1, \ldots x_k)$. One of the most popular techniques used to perform inference is *belief propagation* [77]. While this message passing algorithm can be implemented by networks of spiking neurons [78], a more promising alternative approach, also well suited to model brain-inspired computation, is to use *sampling techniques* [79]. Probably the most important family of sampling techniques in this context is Markov-Chain Monte Carlo (MCMC) sampling. Since MCMC sampling techniques operate in a stochastic manner, stochastic computational elements are a crucial and essential feature. Recent studies have shown that probabilistic inference through MCMC sampling can be implemented by networks of stochastically spiking neurons [79, 80]. Therefore, MCMC sampling is a computational paradigm optimally suited for emulating probabilistic inference in the brain using neuromorphic circuits and nanoelectronic synapses.

Within this context, it is important to see if and how the distribution $P(\bar{x})$ can be *learned* from observations, i.e., how the artificial neural system can build its own model of the world based on its sensory input and then perform probabilistic inference on this model. For a relatively simple model [81], it has been shown that this can be accomplished by a local spike-driven learning rule that resembles the STDP mechanisms measured in cortical networks [50]. Analogous learning mechanisms have been demonstrated both experimentally in neuromorphic CMOS devices [69], and theoretically, with circuit models of memristive synapses [25].

With regard to learning, the variability and stochasticity "features" described above can provide an additional benefit: for many learning tasks, humans and animals have to explore many different actions in order to be able to learn appropriate responses in a given situation. In these so-called reinforcement learning setups, noise and variability naturally provide the required exploration mechanisms. A number of recent studies have shown how stochastic neuronal behavior could be utilized by cortical circuits in order to learn complex tasks [82, 83, 84]. For example, Reservoir Computing (RC, also known under the terms Liquid State Machines and Echo State Networks) is a powerful general principle for computation and learning with complex dynamical systems such as recurrent networks of analog and spiking neurons [85, 86] or optoelectronic devices [87]. The main idea behind RC is to use a heterogeneous dynamical system (called the reservoir) as a nonlinear fading memory where information about previous inputs can be extracted from the current state of the system. This reservoir can be quite arbitrary in terms of implementation and parameter setting as long as it operates in a suitable dynamic regime [88]. Readout elements are trained to extract task-relevant

information from the reservoir. In this way, arbitrary fading memory filters or even arbitrary dynamical systems (in the case when the readout elements provide feedback to the dynamical system) can be learned. One long-standing disadvantage of traditional RC was that readouts had to be trained in a supervised manner. In other words, a teacher signal was necessary that signals at each time point the desired output of readouts. In many real-world applications, such a teacher signal is not available. For example, if the task for a robot controller is to produce some motor trajectory in order to produce a desired hand movement, the exact motor commands that perform this movement are in general not known. What can be evaluated however is the quality of the movement. Recently, it has been demonstrated that noisy readouts can be trained with a much less informative reward signal, which just indicates whether some measure of performance of the system has recently increased [84]. Of course, such reward-based learning can in general be much slower than the pure supervised approach (see, e.g.,[89]). The actual slowdown however depends on the task at hand, and it is interesting that for a set of relevant tasks, reward-based learning works surprisingly fast [84].

Since the functionality of reservoirs depends on its general dynamical behavior and not on precise implementation of its components, RC is an attractive computational paradigm for circuits comprised of nanoscale elements affected by variability, such as the one proposed in Section 4.2. In fact, if the reservoir is composed by a large number of simple interacting dynamic elements – the typical scenario – then heterogeneity of these elements is an essential requirement for ideal performance. Parameter heterogeneity is also beneficial in so-called ensemble learning techniques [90]. It is well-known that the combination of models with heterogeneous predictions for the same data-set tends to improve overall prediction performance [91]. Hence, heterogeneity of computational elements can be a real benefit for learning. Examples for ensemble methods are random forests [92], bagging [93], and boosting [94].

6. Discussion and conclusions

Memristors, and in particular nanoscale solid state implementations, represent a promising technology, baring benefits for emerging memory storage as well as revisiting conventional analog circuits [95]. Given their low-power and small-scale characteristics, researchers are considering their application also in large-scale neural networks for neuro-computing applications. However, the fabrication of large-scale nano-scale cross-bar arrays involves several issues that are still open: the realization of nano sized electrodes requires nanopatterning [96] techniques, such as Electron Beam Lithography (EBL) or Nano-Imprint Lithography (NIL) [97]. This directly correlates to reduced electrode cross section which results in increasing resistance. As electrode resistance scales with length, this can rapidly become a critical issue for fully interconnected nanoscale cross-bar structures. Furthermore, down-scaling the electrode size to reduce the device active area requires simultaneous down-scaling of the thickness of the metalizations due to fabrication concerns. This in turn further increases the resistance of the electrodes,

much like the interconnects in modern CMOS circuitry. These factors introduce a large offset in the write voltages required to change the state of ReRAMs cells that depends on the position of the cell in the array. This problem is especially critical in neuro-computing architectures where these cells represent synapses, as the offsets directly affect the weight-update and learning mechanisms.

Integrating memristors as synapse elements in large-scale neuro-computing architectures also introduces the significance of process variability in memristor dimensions [98], which in turn introduces a significant amount of variability in the characteristics of the synapse properties. In addition to their large variability, another important issue relating to these types of synapses, that is still ignored in the vast majority of neuro-computing studies, is the effect of limited resolution in memristive states. In particular, it is not known what the trade-off between desired synaptic weight resolution and memristor size is. And it is not known to what extent the multi-step synaptic weight model holds true for aggressively down-scaled memristor sizes.

These scaling, integration, and variability issues are serious limiting factors for the use of memristors in conventional neuro-computing architectures. Nonetheless, biological neural systems are an existence proof that it is possible to implement robust computation using nanoscale unreliable components and non-von Neumann computing architectures. In order to best exploit these emerging nanoscale technologies for building compact, low-power, and robust artificial neural processing systems it is important to understand the (probabilistic) neural and cortical principles of computation and to develop at the same time, following a co-design approach, the neuromorphic hardware computing substrates that support them. In this paper we elaborated on this neuromorphic approach, presenting an example of a neuromorphic circuit and of a hybrid nanoelectronic-CMOS architecture that directly emulate the properties of real synapses to reproduce biophysically realistic response properties, thus providing the necessary technology for implementing massively parallel models of brain-inspired computation that are, by design, probabilistic, robust to variability, and fault tolerant.

Acknowledgment

References

[1] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943. 2

[2] J. von Neumann. *The Computer and the Brain*. Yale University Press, New Haven, CT, USA, 1958. 2

[3] F Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review,*, 65(6):386–408, Nov 1958. 2

[4] Marvin L. Minsky. *Computation: finite and infinite machines*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1967. 2

[5] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982. 2

[6] D.E. Rumelhart and J.L. McClelland. *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations.* MIT Press, Cambridge, MA, USA, 1986. 2

[7] T. Kohonen. *Self-Organization and Associative Memory.* Springer Series in Information Sciences. Springer Verlag, 2nd edition, 1988. 2

[8] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Addison-Wesley, Reading, MA, 1991. 2

[9] C.M. Bishop. *Pattern recognition and machine learning.* Springer New York, 2006. 2

[10] Giacomo Indiveri and Timothy K Horiuchi. Frontiers in neuromorphic engineering. *Frontiers in Neuroscience*, 5(118), 2011. 2, 10

[11] C. Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–36, 1990. 2, 10

[12] Henry Markram. The blue brain project. In *ACM/IEEE conference on Supercomputing, SC 2006*, page 53, New York, NY, USA, 2006. IEEE, ACM. 2

[13] J. Schemmel, J. Fieres, and K. Meier. Wafer-scale integration of analog neural networks. In *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 431–438, june 2008. 2

[14] X. Jin, M Lujan, L.A. Plana, S. Davies, S. Temple, and S. Furber. Modeling spiking neural networks on SpiNNaker. *Computing in Science & Engineering*, 12(5):91–97, September-October 2010. 2

[15] J. V. Arthur, P. A. Merolla, F. Akopyan, R. Alvarez, A. Cassidy, A. Chandra, S. K. Esser, N. Imam, W. Risk, D. B. D. Rubin, R. Manohar, and D. S. Modha. Building block of a programmable neuromorphic substrate: A digital neurosynaptic core. In *International Joint Conference on Neural Networks, IJCNN 2012*, pages 1946–1953. IEEE, Jun 2012. 3

[16] T.M. Wong, R. Preissl, P. Datta, M. Flickner, R. Singh, S.K. Esser, E. McQuinn, R. Appuswamy, W.P. Risk, H.D. Simon, and D.S. Modha. 10^{14}. IBM Research Report RJ10502 (ALM1211-004), IBM Research, San Jose, CA, Nov. 2012. 3

[17] R. Silver, K. Boahen, S. Grillner, N. Kopell, and K.L. Olsen. Neurotech for neuroscience: unifying concepts, organizing principles, and emerging tools. *Journal of Neuroscience*, 27(44):11807, 2007. 3

[18] Swadesh Choudhary, Steven Sloan, Sam Fok, Alexander Neckar, Eric Trautmann, Peiran Gao, Terry Stewart, Chris Eliasmith, and Kwabena Boahen. Silicon neurons that compute. In Alessandro Villa, Wlodzislaw Duch, Péter Érdi, Francesco Masulli, and Günther Palm, editors, *Artificial Neural Networks and Machine Learning – ICANN 2012*, volume 7552 of *Lecture Notes in Computer Science*, pages 121–128. Springer Berlin / Heidelberg, 2012. 3

[19] M. Di Ventra and Y.V. Pershin. Memory materials: a unifying description. *Materials Today*, 14(12):584–591, 2011. 3

[20] M Rozenberg, I Inoue, and M Sánchez. Nonvolatile Memory with Multilevel Switching: A Basic Model. *Physical Review Letters*, 92(17):178302, April 2004. 3

[21] B Govoreanu, GS Kar, Y-Y Chen, V Paraschiv, S Kubicek, A Fantini, IP Radu, L Goux, S Clima, R Degraeve, N Jossart, O Richard, T Vandeweyer, K Seo, P Hendrickx, G Pourtois, H Bender, L Altimime, DJ Wouters, JA Kittl, and M Jurczak. $10 \times 10\,nm^2$ Hf/HfOx crossbar resistive RAM with excellent performance, reliability and low-energy operation. *International Technical Digest on Electron Devices Meeting*, pages 31–34, December 2011. 3, 6, 9

[22] Gregory S Snider and R Stanley Williams. Nano/CMOS architectures using a field-programmable nanowire interconnect. *Nanotechnology*, 18(3):035204, January 2007. 4

[23] Tsuyoshi Hasegawa, Takeo Ohno, Kazuya Terabe, Tohru Tsuruoka, Tomonobu Nakayama, James K Gimzewski, and Masakazu Aono. Learning Abilities Achieved by a Single Solid-State Atomic Switch. *Advanced Materials*, 22(16):1831–1834, April 2010. 4

[24] T. Serrano-Gotarredona, T. Masquelier, T. Prodromakis, G. Indiveri, and B. Linares-Barranco. STDP and STDP variations with memristors for spiking neuromorphic learning systems. *Frontiers in Neuroscience*, 7(2), 2013. 4, 7, 8

[25] C. Zamarreño-Ramos, L.A. Camuñas-Mesa, J.A. Pérez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, and B. Linares-Barranco. On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Frontiers in neuroscience*, 5, 2011. 4, 7, 8, 15

[26] Sung Hyun Jo, Ting Chang, Idongesit Ebong, Bhavitavya B Bhadviya, Pinaki Mazumder, and Wei Lu. Nanoscale Memristor Device as Synapse in Neuromorphic Systems. *Nano Letters*, 10(4):1297–1301, April 2010. 4

[27] Hyejung Choi, Heesoo Jung, Joonmyoung Lee, Jaesik Yoon, Jubong Park, Dong-jun Seong, Wootae Lee, Musarrat Hasan, Gun-Young Jung, and Hyunsang Hwang. An electrically modifiable synapse array of resistive switching memory. *Nanotechnology*, 20(34):345201, August 2009. 4

[28] Duygu Kuzum, Rakesh G D Jeyasingh, Byoungil Lee, and H S Philip Wong. Nanoelectronic Programmable Synapses Based on Phase Change Materials for Brain-Inspired Computing. *Nano Letters*, 12(5):2179–2186, May 2012. 4, 7

[29] Leon Chua. Resistance switching memories are memristors. *Applied Physics A*, 102(4):765–783, January 2011. 4

[30] L Chua. Memristor-The missing circuit element. *Circuit Theory, IEEE Transactions on*, 18(5):507–519, 1971. 4

[31] M. Di Ventra, Y.V. Pershin, and L.O. Chua. Circuit elements with memory: memristors, memcapacitors, and meminductors. *Proceedings of the IEEE*, 97(10):1717–1724, 2009. 4

[32] T Prodromakis, C Toumazou, and L Chua. Two centuries of memristors. *Nature Materials*, 11(6):478–481, 2012. 4

[33] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. *Nature*, 459(7250):1154–1154, June 2009. 4

[34] Themistoklis Prodromakis, Boon Pin Peh, Christos Papavassiliou, and Christofer Toumazou. A Versatile Memristor Model With Nonlinear Dopant Kinetics. *Electron Devices, IEEE Transactions on*, 58(9):3099–3105, 2011. 4

[35] Marcus Wu Shihong, Themistoklis Prodromakis, Iulia Salaoru, and Christofer Toumazou. Modelling of Current Percolation Channels in Emerging Resistive Switching Elements. *arXiv.org*, cond-mat.mes-hall, June 2012. 4

[36] Kuk-Hwan Kim, Siddharth Gaba, Dana Wheeler, Jose M Cruz-Albrecht, Tahir Hussain, Narayan Srinivasa, and Wei Lu. A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. *Nano letters*, 12(1):389–395, 2011. 5, 6

[37] Ming Liu, Z Abid, Wei Wang, Xiaoli He, Qi Liu, and Weihua Guan. Multilevel resistive switching with ionic and metallic filaments. *Applied Physics Letters*, 94(23):233106–233106, 2009. 5

[38] Kyung Jean Yoon, Min Hwan Lee, Gun Hwan Kim, Seul Ji Song, Jun Yeong Seok, Sora Han, Jung Ho Yoon, Kyung Min Kim, and Cheol Seong Hwang. Memristive tri-stable resistive switching at ruptured conducting filaments of a pt/tio2/pt cell. *Nanotechnology*, 23(18):185202, 2012. 5

[39] Xiang Yang and I-Wei Chen. Dynamic-load-enabled ultra-low power multiple-state rram devices. *Scientific Reports*, 2, 2012. 5

[40] M.J. O'Donovan and J. Rinzel. Synaptic depression: a dynamic regulator of synaptic communication with varied functional roles. *Trends in Neurosciences*, 20(10):431–3, 1997. 5

[41] F.S. Chance, S.B. Nelson, and L.F. Abbott. Synaptic depression and the temporal response characteristics of V1 cells. *The Journal of Neuroscience*, 18(12):4785–99, 1998. 5

[42] Andras Gelencser, Themistoklis Prodromakis, Christofer Toumazou, and Tamás Roska. A Biomimetic Model of the Outer Plexiform Layer by Incorporating Memristive Devices. *Physical Review E*, 2012. 5

[43] L.M. Yang, Y.L. Song, Y. Liu, Y.L. Wang, X.P. Tian, M. Wang, Y.Y. Lin, R. Huang, Q.T. Zou,

and J.G. Wu. Linear scaling of reset current down to 22-nm node for a novel RRAM. *IEEE Electron Device Letters*, 33(1):89 –91, January 2012. 6

[44] D. Deleruyelle, M. Putero, T. Ouled-Khachroum, M. Bocquet, M.-V. Coulet, X. Boddaert, C. Calmes, and C. Muller. Ge2Sb2Te5 layer used as solid electrolyte in conductive-bridge memory devices fabricated on flexible substrate. *Solid-State Electronics*, 79(0):159–165, January 2013. 6

[45] Albert Chin, Y. C. Chiu, C. H. Cheng, Z. W. Zheng, and Ming Liu. Ultra-low switching power RRAM using hopping conduction mechanism. *Meeting Abstracts*, MA2012-02(31):2574–2574, June 2012. 6

[46] D.L. Lewis and H.-H.S. Lee. Architectural evaluation of 3D stacked RRAM caches. In *IEEE International Conference on 3D System Integration, 2009. 3DIC 2009*, pages 1 –4, September 2009. 6

[47] Dhireesha Kudithipudi and Cory E. Merkel. Reconfigurable memristor fabrics for heterogeneous computing. In Robert Kozma, Robinson E. Pino, and Giovanni E. Pazienza, editors, *Advances in Neuromorphic Memristor Science and Applications*, number 4 in Springer Series in Cognitive and Neural Systems, pages 89–106. Springer Netherlands, January 2012. 7

[48] Konstantin K. Likharev. CrossNets: neuromorphic hybrid CMOS/Nanoelectronic networks. *Science of Advanced Materials*, 3(3):322–331, 2011. 7

[49] G. Indiveri, B. Linares-Barranco, T.J. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. Saighi, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen. Neuromorphic silicon neuron circuits. *Frontiers in Neuroscience*, 5:1–23, 2011. 7, 12

[50] G-Q. Bi and M-M. Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Jour. of Neuroscience*, 18(24):10464–10472, 1998. 7, 15

[51] G.S. Snider. Spike-timing-dependent learning in memristive nanodevices. In *Nanoscale Architectures, 2008. NANOARCH 2008. IEEE International Symposium on*, pages 85–92. IEEE, 2008. 7

[52] B. Linares-Barranco and T. Serrano-Gotarredona. Exploiting memristance in adaptive asynchronous spiking neuromorphic nanotechnology systems. In *Nanotechnology, 2009. IEEE-NANO 2009. 9th IEEE Conference on*, pages 601–604. IEEE, 2009. 8

[53] S.R. Deiss, R.J. Douglas, and A.M. Whatley. A pulse-coded communications infrastructure for neuromorphic systems. In W. Maass and C.M. Bishop, editors, *Pulsed Neural Networks*, chapter 6, pages 157–78. MIT Press, 1998. 8

[54] E. Chicca, A.M. Whatley, P. Lichtsteiner, V. Dante, T. Delbruck, P. Del Giudice, R.J. Douglas, and G. Indiveri. A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity. *IEEE Transactions on Circuits and Systems I*, 5(54):981–993, 2007. 8

[55] K. Likharev, A. Mayr, I. Muckra, and Ö. Türel. Crossnets: High-performance neuromorphic architectures for cmol circuits. *Annals of the New York Academy of Sciences*, 1006(Molecular Electronics III):146–163, 2003. 8

[56] C. Zamarreño-Ramos, A. Linares-Barranco, T. Serrano-Gotarredona, and B. Linares-Barranco. Multi-casting mesh AER: A scalable assembly approach for reconfigurable neuromorphic structured AER systems. application to ConvNets. *Biomedical Circuits and Systems, IEEE Transactions on*, PP(99), 2012. 9

[57] C. Zamarreño-Ramos, T. Serrano-Gotarredona, and B. Linares-Barranco. A 0.35 *murmm* sub-ns wake-up time ON-OFF switchable LVDS driver-receiver chip I/O pad pair for rate-dependent power saving in AER bit-serial links. *Biomedical Circuits and Systems, IEEE Transactions on*, 6(5):486 –497, Oct. 2012. 9

[58] C. Zamarreño-Ramos, T. Serrano-Gotarredona, and B. Linares-Barranco. An instant-startup

jitter-tolerant manchester-encoding serializer/deserializar scheme for event-driven bit-serial LVDS inter-chip AER links. *Circuits and Systems Part-I, IEEE Transactions on*, 58(11):2647–2660, Nov. 2011. 9

[59] C. Tomazou, F.J. Lidgey, and D.G. Haigh, editors. *Analogue IC design: the current-mode approach*. Peregrinus, Stevenage, Herts., UK, 1990. 10

[60] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R.J. Douglas. *Analog VLSI:Circuits and Principles*. MIT Press, 2002. 10, 11

[61] S. Mitra, G. Indiveri, and R. Etienne-Cummings. Synthesis of log-domain integrators for silicon synapses with global parametric control. In *International Symposium on Circuits and Systems, ISCAS 2010*, pages 97–100. IEEE, 2010. 10, 14

[62] C. Bartolozzi and G. Indiveri. Synaptic dynamics in analog VLSI. *Neural Computation*, 19(10):2581–2603, Oct 2007. 11, 12

[63] C. Bartolozzi, S. Mitra, and G. Indiveri. An ultra low power current–mode filter for neuromorphic systems and biomedical signal processing. In *Biomedical Circuits and Systems Conference, BIOCAS 2006*, pages 130–133. IEEE, 2006. 11

[64] B. Gilbert. Translinear circuits: An historical review. *Analog Integrated Circuits and Signal Processing*, 9(2):95–118, March 1996. 11

[65] L.F. Abbott and S.B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, November 2000. 11

[66] G.G. Turrigiano and S.N. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5:97–107, February 2004. 11

[67] R. Brette and W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94:3637–3642, 2005. 12

[68] G. Indiveri and E. Chicca. A VLSI neuromorphic device for implementing spike-based neural networks. In *Neural Nets WIRN11 - Proceedings of the 21st Italian Workshop on Neural Nets*, pages 305–316, Jun 2011. 12

[69] S. Mitra, S. Fusi, and G. Indiveri. Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI. *Biomedical Circuits and Systems, IEEE Transactions on*, 3(1):32–42, Feb. 2009. 12, 13, 15

[70] C. Bartolozzi and G. Indiveri. Global scaling of synaptic efficacy: Homeostasis in silicon synapses. *Neurocomputing*, 72(4–6):726–731, Jan 2009. 12

[71] J. Brader, W. Senn, and S. Fusi. Learning real world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, 19:2881–2912, 2007. 13

[72] E. Todorov and M. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(1):1226–1235, 2002. 14

[73] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008. 14

[74] M. Volgushev, I. Kudryashov, M. Chistiakova, M. Mukovski, J. Niesmann, and U.T. Eysel. Probability of transmitter release at neocortical synapses at different temperatures. *Journal of neurophysiology*, 92(1):212–220, 2004. 14

[75] T. L. Griffiths and J. B. Tenenbaum. Optimal predictions in everyday cognition. *Psychological Science*, 17(9):767–773, 2006. 14

[76] Joshua B. Tenenbaum, Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011. 14

[77] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988. 14, 15

[78] A. Steimer, W. Maass, and R. Douglas. Belief propagation in networks of spiking neurons. *Neural Comput*, 21(9):2502–2523, Sep 2009. 15

[79] L. Büsing, J. Bill, B. Nessler, and W. Maass. Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, 7(11):e1002211, 2011. 15

[80] D. Pecevski, L. Büsing, and W. Maass. Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. *PLoS Computational Biology*, 7(12):e1002294, 2011. 15

[81] B. Nessler, M. Pfeiffer, and W. Maass. STDP enables spiking neurons to detect hidden causes of their inputs. In *Proc. of NIPS 2009: Advances in Neural Information Processing Systems*, volume 22, pages 1357–1365. MIT Press, 2010. 15

[82] R. Legenstein, D. Pecevski, and W. Maass. A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLoS Computational Biology*, 4(10):e1000180, 2008. 15

[83] R. Legenstein, S. M. Chase, A. B. Schwartz, and W. Maass. A reward-modulated Hebbian learning rule can explain experimentally observed network reorganization in a brain control task. *The Journal of Neuroscience*, 30(25):8400–8410, 2010. 15

[84] G. M. Hoerzer, R. Legenstein, and Wolfgang Maass. Emergence of complex computational structures from chaotic neural networks through reward-modulated Hebbian learning. *Cerebral Cortex*, 2012. in press. 15, 16

[85] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless telecommunication. *Science*, 304(5667):78–80, April 2 2004. 15

[86] W. Maass, T. Natschlaeger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002. 15

[87] Yvan Paquot, Francois Duport, Antoneo Smerieri, Joni Dambre, Benjamin Schrauwen, Marc Haelterman, and Serge Massar. Optoelectronic reservoir computing. *SCIENTIFIC REPORTS*, 2:1–6, 2012. 15

[88] R. Legenstein and W. Maass. *What makes a dynamical system computationally powerful?*, pages 127–154. MIT Press, 2007. 15

[89] Robert Urbanczik and Walter Senn. Reinforcement learning in populations of spiking neurons. *Nature neuroscience*, 12(3):250–252, 2009. 16

[90] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33:1–39, 2010. 16

[91] Ludmila I. Kuncheva and Christopher J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.*, 51(2):181–207, May 2003. 16

[92] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. 16

[93] Leo Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996. 16

[94] Robert E. Schapire. The strength of weak learnability. *Mach. Learn.*, 5(2):197–227, July 1990. 16

[95] R. Berdan, T. Prodromakis, I. Salaoru, A. Khiat, and C. Toumazou. Memristive devices as parameter setting elements in programmable gain amplifiers. *Applied Physics Letters*, 101(24):243502–243502, 2012. 16

[96] Qiangfei Xia, J. Joshua Yang, Wei Wu, Xuema Li, and R. Stanley Williams. Self-aligned memristor cross-point arrays fabricated with one nanoimprint lithography step. *Nano Letters*, 10(8):2909–2914, August 2010. 16

[97] Qiangfei Xia. Nanoscale resistive switches: devices, fabrication and integration. *Applied Physics A*, 102(4):955–965, March 2011. 16

[98] Miao Hu, Hai Li, Yiran Chen, Xiaobin Wang, and Robinson E. Pino. Geometry variations analysis of TiO_2 thin-film and spintronic memristors. In *Proceedings of the 16th Asia and South Pacific Design Automation Conference*, ASPDAC '11, pages 25–30, Piscataway, NJ, USA, 2011. IEEE Press. 17