Artificial Cognitive Systems

From VLSI Networks of Spiking Neurons to Neuromorphic Cognition

Cogn Comput (2009) 1:119–127
DOI 10.1007/s12559-008-9003-6

Artificial Cognitive Systems: From VLSI Networks of Spiking Neurons to Neuromorphic Cognition

Giacomo Indiveri · Elisabetta Chicca · Rodney J. Douglas

Published online: 14 January 2009

Abstract Neuromorphic engineering (NE) is an emerging research field that has been attempting to identify neural types of computational principles, by implementing biophysically realistic models of neural systems in Very Large Scale Integration (VLSI) technology. Remarkable progress has been made recently, and complex artificial neural sensory-motor systems can be built using this technology. Today, however, NE stands before a large conceptual challenge that must be met before there will be significant progress toward an age of genuinely intelligent neuromorphic machines. The challenge is to bridge the gap from reactive systems to ones that are cognitive in quality. In this paper, we describe recent advancements in NE, and present examples of neuromorphic circuits that can be used as tools to address this challenge. Specifically, we show how VLSI networks of spiking neurons with spike-based plasticity mechanisms and soft winner-take-all architectures represent important building blocks useful for implementing artificial neural systems able to exhibit basic cognitive abilities.

G. Indiveri · E. Chicca · R. J. Douglas
Institute of Neuroinformatics, University of Zurich,
ETH Zurich, Zurich, Switzerland

E. Chicca
e-mail: chicca@ini.phys.ethz.ch

R. J. Douglas
e-mail: rjd@ini.phys.ethz.ch

G. Indiveri (✉) · E. Chicca · R. J. Douglas
University of Zurich, ETH Zurich, Zurich, Switzerland
e-mail: giacomo@ini.phys.ethz.ch

Keywords Neuromorphic engineering · Cognition · Spike-based learning · Winner-take-all · Soft WTA · VLSI

Introduction

Machine simulation of cognitive functions has been a challenging research field since the advent of digital computers. Despite the resources dedicated to this field, humans, mammals, and many other animal species, including insects, still outperform the most powerful computers in relatively routine functions such as, for example, vision. The disparity between the effectiveness of computation in biological nervous systems and in a computer, in such types of functions, is primarily attributable to the way the elementary devices are used in the system, and to the kind of computational primitives they implement [48]. Rather than using Boolean logic, precise digital representations, and clocked operations, nervous systems carry out robust and reliable computation using hybrid analog/digital unreliable components; they emphasize distributed, event-driven, collective, and massively parallel mechanisms, and make extensive use of adaptation, self-organization, and learning. Understanding these principles, and how they can lead to behaviors that exhibit cognitive qualities is one of the major challenges of modern science.

Neuromorphic engineering (NE) is a research field that is addressing these issues by designing and fabricating electronic neural systems whose architecture and design principles are based on those of biological nervous systems. The term neuromorphic engineering was coined by Carver Mead in the late 1980s to describe Very Large Scale Integration (VLSI) systems comprising analog circuits and built to mimic biological neural cells and architectures

[47]. Since then, NE has been attempting to identify neural types of computational principles, by implementing biophysically realistic models of neural systems in VLSI technology to reproduce the same physics of computation [27].

During the last decade, the NE community has made substantial progress by designing hybrid analog/digital circuits that implement silicon neurons and synapses [7, 38, 66], silicon retinas, and cochleas [13, 43], and by developing the technology for constructing distributed multichip systems of sensors and neuronal processors that operate asynchronously and communicate using action-potential-like signals (or spikes) [17, 49]. A number of research groups worldwide are already developing large scale (in terms of component count) neuromorphic systems [59, 60]. Today, however, NE stands before a large conceptual challenge that must be met before there will be significant progress toward an age of genuinely intelligent neuromorphic machines. The challenge is to bridge the gap from reactive systems to ones that are cognitive in quality. In NE, as in neuroscience and computer science, we understand very little about how to configure these large systems to achieve the sophistication of processing that we could regard as effective *cognition*.

In the case of NE and neuroscience, the question is sharpened by the need to understand cognition in the context of the nervous systems' peculiar hardware and style of processing. We know, for instance, that nervous systems can exhibit context-dependent behavior, can execute "programs" consisting of series of flexible steps, and can conditionally branch to alternative behaviors, using spiking neurons and dynamic synapses as basic computational modules.

The NE community has recently developed efficient VLSI implementations of such types of computational modules: next to several designs of conductance-based and integrate-and-fire neurons [19, 25, 38, 58, 66], NE researchers proposed circuits that implement VLSI dynamic synapses [7], spike-based plasticity mechanisms [32, 34, 50, 68], and soft winner-take-all (WTA) networks [16], for example.VLSI implementations of WTA networks of spiking neurons, with plastic dynamic synapse circuits are particularly important, because recent theoretical studies demonstrated that recurrent neural networks arranged in a way to implement soft WTA performance can implement critical aspects of cortical computation [57].

In the next section, we present an overview of the recent advances made in neuromorphic VLSI circuit design of spiking neural networks, soft WTA networks, and spike-based plasticity mechanisms. While in the "Neuromorphic Cognition" section, we describe the "neuromorphic cognition" challenge, arguing that VLSI networks of spiking neurons with spike-based plasticity mechanisms and soft WTA architectures represent a crucial building block useful for constructing future VLSI neuromorphic cognitive systems.

Neuromorphic VLSI

When implemented in VLSI technology, neuromorphic circuits use, to a large extent, the same physics used in neural systems (e.g., they transport majority carriers across the channel of transistors by diffusion processes, very much like neurons transport ions inside or outside cell bodies through their proteic channels). Given the analogies at the single device level, larger scale neuromorphic circuits share many common physical constraints with their biological counterparts (given by noise, temperature dependence, inhomogeneities, etc.). Therefore, these architectures often have to use similar strategies for carrying out computation while maximizing compactness, optimizing robustness to noise, minimizing power consumption, and increasing fault tolerance.

In recent years, an interesting class of neuromorphic devices implementing general-purpose computational architectures based on networks of silicon neurons and synapses started to emerge. These devices range from reconfigurable arrays of basic integrate and fire neuron models [17, 18, 38, 45, 48], to learning architectures implementing detailed models of spike-based synaptic plasticity [5, 6, 38, 50, 53, 56]. Spike-based plasticity circuits enable these systems to adapt to the statistics of their input signals, to learn and classify complex sequences of spatio-temporal patterns (e.g., arising from visual or auditory signals), and eventually to interact with the user and the environment.

Consistent with the NE approach, the strategy used to transmit signals across chip boundaries in these types of systems is inspired from the nervous system: output signals are represented by stereotyped digital pulses (spikes), and the analog nature of the signal is typically encoded in the mean frequency of the neuron's pulse sequence (spike rates) and the instantaneous inter-spike interval (ISI). Similarly, input signals are represented by spike trains, conveyed to the chip in the form of asynchronous digital pulses, that stimulate their target synapses on the receiving chip. The circuits that generate the on-chip synaptic currents when stimulated by incoming spikes are slow low-power analog circuits. The circuits that generate and manage these streams of input/output digital pulses are fast asynchronous logic elements, based on an emerging new communication standard for neuromorphic chips called the "address-event representation" (AER) [17, 21, 42].

By using both low-power analog circuits and self-clocked asynchronous digital logic neuromorphic devices take advantage of the best of both worlds. Using a real-time

asynchronous digital communication infrastructure, arbitrarily complex systems can be constructed by interfacing multiple chips together. Substantial technological advancements have been made in this domain, and the VLSI design aspect of these devices has reached a mature state. However, although it is now clear how to implement large distributed networks of spiking neurons with plastic synapses distributed across multiple chips, there have been no systematic attempts so far to use this technology for modeling cognitive processes. Neither has there been a systematic study so far to determine how to implement cognitive behaviors with networks of spiking neurons, which can be directly mapped onto multi-chip networks of silicon neurons. Similarly, there have been very few attempts at using neuromorphic networks of spiking neurons on robotic platforms, for implementing real-time spike-based learning, adaptation, and context-dependent action selection, for example, in behaving systems.

Soft Winner-Take-All Circuits

Winner-take-all networks of spiking neurons are ideally suited for implementing context-dependent action selection operators. These types of networks typically consist of a group of interacting neurons which compete with each other in response to an input stimulus. The neurons with highest response suppress all other neurons to win the competition. Competition is achieved through a recurrent pattern of connectivity involving both excitatory and inhibitory connections. Cooperation between neurons with similar response properties (e.g., close receptive field or stimulus preference) is mediated by excitatory connections. Competition and cooperation make the output of individual neuron depend on the activity of all neurons in the network and not just on its own input. As a result, these networks performs not only common linear operations but also complex nonlinear operations (see Fig. 1). The linear operations include analog gain and locus invariance [36]. The nonlinear operations include nonlinear selection or *soft winner-take-all* behavior [3, 20, 67], signal restoration [20, 26], and multi-stability [3, 35, 67].

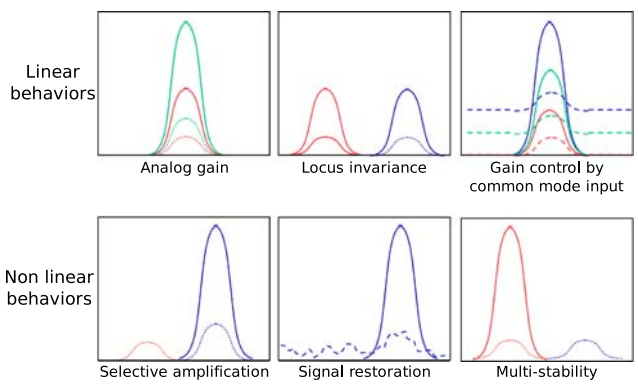

Fig. 1 Linear and nonlinear behaviors expressed by soft WTA networks

The computational abilities of soft WTA networks have been used for solving feature extraction and pattern classification problems [8, 9, 61]. When soft WTA networks are used for solving classification tasks, common features of the input space can be learned in an unsupervised manner. For example, Bennett [9] showed that competition supports unsupervised learning because it enhances the firing rate of the most excited neurons (i.e., the ones receiving the strongest input) which, in turn, triggers learning.

Soft WTA networks are believed to play a central role in cortical processing. A majority of synapses in the mammalian cortex originate within the cortex itself [10, 28]. Neurons with similar functional properties are aggregated together in *modules* or *columns* and most connections are made locally within the neighborhood of a 1 mm column [41]. Soft WTA models try to emulate the cortical pattern of connectivity and to study its role in processing sensory inputs and in generating behavioral outputs.

The highly distributed nature of physical computation in these types of neural networks can be faithfully reproduced using neuromorphic circuits that implement networks of integrate-and-fire neurons and plastic synapses in VLSI technology.

Several examples of VLSI WTA networks of spiking neurons can be found in the literature [2, 15, 24, 37, 40, 51]. In 1992, De Yong et al. [24] proposed a VLSI WTA spiking network consisting of four neurons. The authors implemented the WTA mechanism through all-to-all inhibitory connections. They showed how their network exhibits two different behaviors depending on the time constant of the inhibitory post-synaptic potential (IPSP) relative to the time period of the incoming signal: (1) the network acts as a *temporal* WTA (only the first neuron receiving an input spike becomes active and wins the competition) when the time constant of the IPSP is longer than the period of the slowest input signal; (2) the network behaves as a maximum *frequency* operator (only the neuron receiving the train of spikes with highest frequency becomes active) when the period of the fastest input signal is longer than the time constant of the IPSP. In both cases, the network behaves as a hard WTA, allowing only one neuron to be active.

In 1993, a three-neuron VLSI WTA chip was proposed by Hylander et al. [37]. Their network used global inhibition to implement the WTA behavior. The three neurons fed their outputs to the global inhibitory generator, which fed back inhibition to all the neurons in the network. Also this network behaved as a hard WTA.

Both De Young et al. and Hylander et al. presented very simple examples of WTA networks, and showed the ability of their VLSI networks to select one winner, but not to perform soft WTA computation. Thanks to the progress of VLSI technology, more recent implementation of spiking VLSI WTA networks integrate many more neurons on a single chip, and implement more elaborate soft WTA models: in 2001, Indiveri et al. [40] presented a spiking network consisting of 32 excitatory neurons and one global inhibitory neuron. The authors characterized the behavior of the network using the mean rate representation and Poisson distributed input spike trains. They showed the network could exhibit soft WTA behaviors modulated by the strength of lateral excitation and investigated the network's ability to produce correlated firing, combined with the WTA function. In 2004, several additional VLSI implementations of WTA networks were presented: Oster and Liu [51] presented a 64 neurons network that used all-to-all inhibition to implement a hard WTA behavior; Abrahamsen et al. [2] presented a time domain WTA network that used self-resetting I&F neurons to implement hard WTA behavior, by resetting all neurons in the array simultaneously, as soon as the winning neuron fired; and Chicca et al. [15] presented a recurrent network of spiking neurons, comprising 31 excitatory neurons and 1 global inhibitory neuron. This network is an evolution of the one presented in [40] which includes second neighbor excitatory connections (in addition to first neighbor excitation), and can be operated in open-(linear array) or closed-loop (ring) conditions. Figure 2 shows experimental data measured from the chip, describing how it is able to perform nonlinear selection, one of the typical soft WTA network behaviors (see also Fig. 1). An input stimulus (see Fig. 2a) consisting of Poisson trains of spikes, with a mean frequency profile showing two Gaussian-shaped bumps with different amplitude, is applied to the input synapses of each neuron in the soft WTA network. The chip output response is a series of spike trains produced by the 32 silicon neurons (see Fig. 2b). The mean frequencies measured from each spike raster in Fig. 2b show how the soft WTA network (blue line) selects and amplifies the Gaussian bump with higher activity while suppressing the other one, with respect to the baseline condition (no recurrent connections, green line).

More recent hardware implementations of the spiking soft WTA network have been realized by the authors. These chips comprise both larger numbers of neurons (e.g., up to 2048) and spike-based learning capabilities (see "Spike-Based Learning" section).

Spike-Based Learning

An additional feature that is crucial for implementing cognitive systems with networks of spiking neurons is *spike-based plasticity*. Plasticity is one of the key properties of biological synapses, which provides the brain with the ability to learn and to form memories. In particular, *long-term plasticity* (LTP) is a mechanism which produces activity-dependent long-term changes in the synaptic strength of individual synapses, and plays a crucial role in learning [1]. A popular class of LTP spike-driven learning mechanisms, that has recently been the subject of widespread interest, is the one based on spike-timing dependent plasticity (STDP) [1, 46]. In STDP, the relative timing of pre- and post-synaptic spikes determine how to update the efficacy of a synapse. In VLSI networks of spiking neurons, STDP-type mechanisms map very effectively onto silicon. Several examples of STDP learning chips have

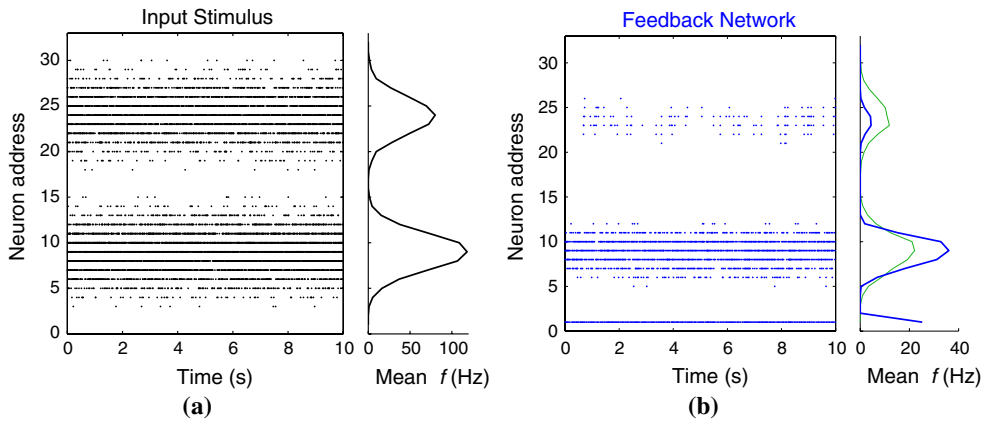

Fig. 2 Raster plot and mean frequency profile of input stimulus (**a**) and network response (**b**). The input stimulus (**a**) consists of Poisson trains of spike, the mean frequency profile over neuron address shows two Gaussian-shaped bumps of activity with different amplitude. **b** The soft WTA network response shows how the bump with higher amplitude is selected and amplified while the other one is suppressed. The response of the feed-forward network (no recurrent connections) to the input, is shown for comparison (green curve in the mean frequency profile)

been recently proposed [5, 38, 54, 68], and it has been shown, both in theoretical models and VLSI implementations, that STDP can be effective in learning to classify spatio-temporal spike patterns [5, 33, 34].

However, when considering learning in physical implementations of synapses, either biological or electronic, a crucial problem arises: their synaptic weights are bounded (i.e., they cannot grow indefinitely or assume negative values) and have limited precision. This constraint, often ignored in software simulations, poses strong limitations on the network's capacity to preserve memories stored in the synaptic weights: if synapses cannot be modified with an arbitrarily large precision, the storage of new memories can overwrite old ones, eventually making memory retrieval (and learning) impossible. In these conditions, the synapses tend to reach a uniform equilibrium distribution very rapidly, at a rate which depends on the extent by which the synapses are modified [4, 29]. A large number of synaptic modifications implies fast learning, but also fast forgetting. Extending the range in which the synapses vary, or their resolution (i.e., the number of discrete stable states that exist between from their lower to their upper bound) does not improve the memory performance considerably [30]. But the memory lifetime can be greatly increased by slowing down the learning process (e.g., by modifying only a small subset of synapses) [4, 29].

A spike-based learning algorithm that takes into account these considerations has been recently proposed in [11]. We developed a hardware implementation of this model using spike-based plasticity circuits with the minimal number of stable states (two), and with a weight-update scheme that consolidates the transitions between one stable state to the other in a stochastic way, to be able to change the weights with a small probability [39]. Using just two stable synaptic states solves efficiently the problem of long-term storage: it is sufficient to use a bistable circuit that restores the synaptic state to either its high rail or its low one, depending on whether the weight is above or below a set threshold. In this way, memory preservation is guaranteed also in the absence of stimuli, or when the pre-synaptic activity is very low. The synaptic weight updated depends on the timing of the pre-synaptic spike, on the state of the post-synaptic neuron's membrane potential, and on a slow variable proportional to the post-synaptic neuron's mean firing rate (related to the Calcium concentration in real neurons). Such a model has been shown to be able to classify patterns of mean firing rates, to capture the rich phenomenology observed in neurophysiological experiments on synaptic plasticity, and to reproduce the classical STDP phenomenology [11].

This particular strategy for spike-based learning is effective for VLSI devices which implement networks of silicon neurons with a large number of bistable synapses, and which can make use of a stochastic mechanism for updating the synaptic weights. Indeed, by modifying only a random subset of all the stimulated synapses with a small probability, the network's memory lifetime increases significantly (memory lifetimes increase by a factor inversely proportional to the probability of synaptic modification) [29]. The stochastic mechanism required for making a random selection of synapses is implemented directly, without the need of special additional circuits such as random-number generators, exploiting the properties of the AER communication protocol. Indeed, if the trains of spikes (address-events) transmitted to the plastic synapse have a Poisson distribution (as is the case for address-events produced by silicon neurons embedded in a recurrent network with sparse connectivity [14, 65]), and the synaptic transition between the two stable states occur only after a sufficient number of spike-driven events accumulate, then the changes in the synaptic weight are stochastic [14, 31].

To validate the VLSI implementation of the learning model proposed in [11], we fabricated a small 10 mm^2 prototype chip comprising an array of 128 integrate-and-fire neurons and 4096 adaptive synapses with biologically plausible temporal dynamics [7], using a standard 0.35 μm CMOS technology (see Fig. 3). We presented experimental data from the chip describing the detailed behavior of the learning circuits in [50], and showed how such circuits can robustly classify complex patterns of spike trains.

The array of neurons implemented in this chip comprises also additional local excitatory and inhibitory synapses to form a soft WTA architecture. Therefore this device, thanks to its spike-based plasticity and soft WTA mechanisms, can be used in distributed multi-chip AER systems as a general purpose computational module, and

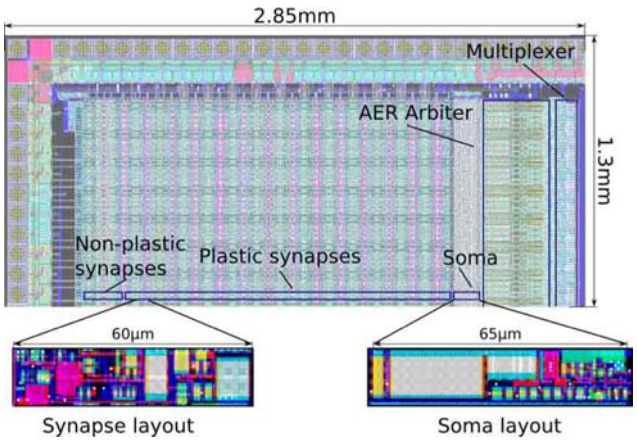

Fig. 3 Layout section of the spike-based learning chip: an array of 128 I&F neurons, represented by the "Soma" block, is connected to 128 rows of 32 AER synapses (28 with plasticity, and 4 nonplastic synapses with fixed weights). An on-chip multiplexer allows the user to select how many rows of synapses/neuron to configure. The AER arbiter is a communication block which transmits the asynchronous address-events off-chip

represents a useful tool for the realization of neuromorphic cognitive systems.

Neuromorphic Cognition

So far the NE endeavor has succeeded in providing the physical infrastructure for constructing networks of sensors, neuronal networks, and effectors that are similar in organization if not in size to the nervous systems of biology. However, the tasks that these neuromorphic systems are able to perform remain rather simple, so that even the most sophisticated VLSI neuromorphic systems created thus far are reactive in quality, mapping rather directly sensory percepts to simple actions. Of course, intelligent systems are not simply reactive. Instead, given some knowledge of its environment and some approximate behavioral objectives, an intelligent agent comes to reason that certain combinations of actions are more likely to achieve an objective than others.

The fact that neuromorphic systems still fall short of such performance is not due to any particular restriction of the available hardware, which has grown evermore sophisticated and reliable. It simply reflects the state of progress of the field. It is only during recent years that it has become technically possible to consider how neuromorphic systems could be configured to perform behavior that is more elaborate than reactive: to consider how to make these systems more cognitive. That the community recognizes this challenge can be seen in the recent establishment in Europe of the vigorous Capo Caccia Workshops toward Cognitive Neuromorphic Engineering [12]; in the redirection of the NSF Telluride Workshops [64] also toward that goal; and in the launch of the DARPA SyNAPSE initiative [63].

The step from reaction to cognition is not an easy one. For a system to exhibit cognition, it must be capable of creating, storing and manipulating knowledge of the world and of itself, and of reasoning on this knowledge to plan and execute economically advantageous behavior. Whereas we may recognize these properties in the behavior of animals, it has been extraordinarily difficult to evoke them in artificial systems, be they either symbolic or connectionist in design. Nor has either Neuroscience or Psychology been quick to identify the organization principles of the brain or mind, which support cognition. By "identify" we mean not simply a description of what there is, but rather an explanation of how things work, in a manner that can be used to develop a practical technology. To the extent that science has been able to evoke artificial cognition at all, it has been based largely on symbolic encodings of the world processed on conventional digital computers that use predominantly nonreal time, serial, synchronized electronic processing. In these systems, the algorithmic processing of information occurs without any intrinsic regard for the meaning and significance of the processed data. The meaning and significance are extrinsic to the computation. They are derived from the interpretation of human programmers who design the encodings of the data, and the algorithms that manipulate them. And so cognition is not inherent to this style of computation. On the other hand, there is no reason to believe that present methods of computation are unable to express cognition. It is likely that intelligence expressed by cognition is a particular style of computation in which the attribution of meaning, significance, purpose, etc. arise out of the self-organization of encodings by the algorithms themselves, rather than the external programmers. The challenge for NE is to establish whether neuromorphic architectures and computation offer any advantage over conventional digital methods for implementing this style of computation. The challenge is not simply about hardware implementation, but more generally to understand what kinds of computational models neurons can support, and how to configure the hardware neurons to perform desired tasks using a particular computational approach.

For example, one fundamental problem is how nervous systems transform the world into a form suitable for the expression of cognition. This is a transformation of the sensory data into a kind of symbolic representation that can support reasoning. Biological sensors use real-valued signals, that must be extracted from noise, amplified, segmented, and combined to form the objects and their relations that are the meat of behavioral action. The sensory data are often incomplete, and must be combined with incomplete theories of how the world functions. How are these theories derived from the world and implemented in neurons? Indeed, the ability to infer unknown information from incomplete sensory data combined with some prior knowledge must rank as one of the most fundamental principles for incorporation in neuronal circuits. Already, there exists interesting progress in this direction. Several recent studies have considered how single neurons or their networks could implement belief propagation [52, 55], or of how they could perform probabilistic computations in general [22, 23, 44, 69]. Steimer et al. [62] have shown how pools of spiking neurons can be used to implement the Belief-Propagation algorithm on a factor graph. The pools of neurons implement the nodes of a factor graph. Each pool gathers 'messages' in the form of population activities from its input nodes and combines them through its network dynamics. The various output messages to be transmitted over the edges of the graph are each computed by a group of readout neurons that feed into their respective destination pools. They use this approach to demonstrate how pools of spiking neurons can explain how visual cues

resolve competing interpretations of an object's shape and illumination. Work such as this shows how networks of neurons can support a rather general computational model (in this case, factor graphs) and how the operation of neurons can be linked to psycho-physical experience.

Another interesting problem is how neurons support conditional branching between possible behavioral states, which is a hallmark of intelligent behavior. In a step toward solving this problem, Rutishauser and Douglas [57] have recently shown how neuronal networks with a nearly uniform architecture can be configured to provide conditional branching between neuronal states. They show that a multistable neuronal network containing a number of states can be created very simply, by coupling two recurrent networks whose synaptic weights have been configured for soft WTA performance. The two soft WTAs have simple, homogeneous locally recurrent connectivity except for a small fraction of recurrent cross-connections between them, which are used to embed the required states. The coupling between the maps allows the network to continue to express the current state even after the input that evoked that state is withdrawn. In addition, a small number of "transition neurons" implement the necessary input-driven transitions between the embedded states. Simple rules are provided to systematically design and construct neuronal state machines of this kind. The significance of this finding is that it offers a method whereby cortex-like plates of neurons could be configured for sophisticated processing by applying only small specializations to the same generic neuronal circuit.

These two examples represent only demonstrations of principle, validated by software simulations. However, they are sufficiently simple in concept and small in network size to be directly implemented in neuromorphic VLSI. The resulting systems, comprising soft WTA neural circuits and plastic synapses previously described, will be useful for exploring more sophisticated neuromorphic behavior.

Conclusions

Neuromorphic engineering has been very successful in developing a new generation of computing technologies implemented with design principles based on those of the nervous systems, and which exploit the physics of computation used in biological neural systems. We are now able to design and implement complex large-scale artificial neural systems with elaborate computational properties, such as spike-based plasticity and soft WTA behavior. It is even possible to build complete artificial sensory-motor systems, able to robustly process signals in real-time using neuromorphic VLSI technology. However, there is still a large gap between the type of *reactive* systems that have been built up to now, and neuromorphic behaving systems able to achieve the sophistication of processing that we could regard as effective *cognition*.

In this paper, we presented an overview of the recent advances made in neuromorphic VLSI technology, focusing on soft WTA networks of spiking neurons and spike-based plasticity mechanisms, and described some of the challenges that the research community faces for bridging this gap and going from NE to neuromorphic cognition. We argued that the silicon neuron and spike-based plasticity circuits discussed in "Neuromorphic VLSI" section can be used to learn to infer unknown information from incomplete sensory data (i.e., implement Belief-Propagation networks), while the soft WTA networks represent a useful computational paradigm for "programming" networks of spiking neurons, thanks also to their ability to implement conditional branching between neuronal states.

The neural network examples that implement Belief-Propagation networks and soft WTA architectures that exhibit conditional branching between neuronal states have only been tested in software models for now, but they can be directly mapped onto neuromorphic multi-chip architectures.

By combining research on neuromorphic VLSI technology, software models of spiking neural architectures, and neuroscience, it will be soon possible to implement artificial systems comprising VLSI networks of spiking neurons, able to exhibit context-dependent cognitive abilities in real-time, and in response to real-world stimuli.

Acknowledgments This work was supported by the DAISY (FP6-2005-015803) EU Grant, by the Swiss National Science Foundation under Grant PMPD2-110298/1, and by the Swiss Federal Institute of Technology Zurich Grant TH02017404.

References

1. Abbott L, Nelson S. Synaptic plasticity: taming the beast. Nat Neurosci 2000;3:1178–83.
2. Abrahamsen J, Hafliger P, Lande T. A time domain winner-take-all network of integrate-and-fire neurons. In: 2004 IEEE international symposium on circuits and systems, vol. 5. 2004. p. V-361–4.
3. Amari S, Arbib MA. Competition and cooperation in neural nets. In: Metzler J, editor. Systems neuroscience. Academic Press; 1977. p. 119–65
4. Amit DJ, Fusi S. Dynamic learning in neural networks with material synapses. Neural Comput. 1994;6:957.
5. Arthur J, Boahen K. Learning in silicon: timing is everything. In: Weiss Y, Schölkopf B, Platt J, editors. Advances in neural information processing systems, vol. 18. Cambridge, MA: MIT Press; 2006. p. 1–8.
6. Badoni D, Giulioni M, Dante V, Del Giudice P. An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses. In: Proceedings of the IEEE international symposium on circuits and systems. IEEE; 2006. p. 1227–30.
7. Bartolozzi C, Indiveri G. Synaptic dynamics in analog VLSI. Neural Comput. 2007;19(10):2581–603.

8. Ben-Yishai R, Lev Bar-Or R, Sompolinsky H. Theory of orientation tuning in visual cortex. Proc Natl Acad Sci USA. 1995;92(9):3844–8.
9. Bennett A. Large competitive networks. Network. 1990;1:449–62.
10. Binzegger T, Douglas RJ, Martin K. A quantitative map of the circuit of cat primary visual cortex. J Neurosci. 2004;24(39):8441–53.
11. Brader J, Senn W, Fusi S. Learning real world stimuli in a neural network with spike-driven synaptic dynamics. Neural Comput. 2007;19:2881–912.
12. The Capo Caccia, workshops toward cognitive neuromorphic engineering. http://cne.ini.uzh.ch/capocaccia08. April 2008.
13. Chan V, Liu SC, van Schaik A. AER EAR: a matched silicon cochlea pair with address event representation interface. IEEE Trans Circuit Syst I. 2006;54(1):48–59, Special issue on sensors.
14. Chicca E, Fusi S. Stochastic synaptic plasticity in deterministic a VLSI networks of spiking neurons. In: Rattay F, editor. Proceedings of the world congress on neuroinformatics, ARGESIM reports, 2001. Vienna: ARGESIM/ASIM Verlag; 2001. p. 468–77.
15. Chicca E, Indiveri G, Douglas R. An event based VLSI network of integrate-and-fire neurons. In: Proceedings of IEEE international symposium on circuits and systems. IEEE; 2004. p. V-357–60.
16. Chicca E, Indiveri G, Douglas R. Context dependent amplification of both rate and event-correlation in a VLSI network of spiking neurons. In: Schölkopf B, Platt J, Hofmann T, editors. Advances in neural information processing systems, vol. 19. Neural Information Processing Systems Foundation. Cambridge, MA: MIT Press; 2007, in press.
17. Chicca E, Whatley AM, Dante V, Lichtsteiner P, Delbrück T, Del Giudice P, et al. A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity. IEEE Trans Circuit Syst I Regular Paper. 2007;5(54):981–93.
18. Choi TYW, Merolla PA, Arthur JV, Boahen KA, Shi BE. Neuromorphic implementation of orientation hypercolumns. IEEE Trans Circuit Syst I. 2005;52(6):1049–60.
19. Culurciello E, Etienne-Cummings R, Boahen K. Arbitrated address-event representation digital image sensor. Electron Lett. 2001;37(24):1443–5.
20. Dayan P, Abbott L. Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT Press; 2001.
21. Deiss S, Douglas R, Whatley A. A pulse-coded communications infrastructure for neuromorphic systems. In: Maass W, Bishop CM. editors. Pulsed neural networks, chapter 6. Cambridge, MA: MIT Press; 1998. p. 157–78.
22. Deneve S. Bayesian spiking neurons 1: inference. Neural Comput. 2007;20(1):91–117.
23. Deneve S, Latham P, Pouget A. Efficient computation and cue integration with noisy population codes. Nat Neurosci. 2001;4(8):826–31.
24. DeYong MR, Findley RL, Fields C. The design, fabrication, and test of a new VLSI hybrid analog-digital neural processing element. IEEE Trans Neural Netw. 1992;3(3):363–74.
25. Douglas R, Mahowald M. Silicon neurons. In: M. Arbib, editor. The handbook of brain theory and neural networks. Boston, MA: MIT Press; 1995. p. 282–9.
26. Douglas R, Mahowald M, Martin K. Hybrid analog-digital architectures for neuromorphic systems. In: Proceedings of IEEE world congress on computational intelligence, vol. 3. IEEE; 1994. p. 1848–53.
27. Douglas R, Mahowald M, Mead C. Neuromorphic analogue VLSI. Annu Rev Neurosci. 1995;18:255–81.
28. Douglas R, Martin K. Neural circuits of the neocortex. Annl Rev Neurosci. 2004;27:419–51.
29. Fusi S. Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. Biol Cybernet. 2002;87:459–70.
30. Fusi S, Abbott LF. Limits on the memory storage capacity of bounded synapses. Nat Neurosci. 2007;10:485–93.
31. Fusi S, Annunziato M, Badoni D, Salamon A, Amit DJ. Spike–driven synaptic plasticity: theory, simulation, VLSI implementation. Neural Comput. 2000;12:2227–58.
32. Giulioni M, Camilleri P, Dante V, Badoni D, Indiveri G, Braun J, et al. A VLSI network of spiking neurons with plastic fully configurable "stop-learning" synapses. In: Proceedings of IEEE international conference on electronics, circuits, and systems, ICECS 2008. IEEE; 2008. p. 678–81.
33. Gütig R, Sompolinsky H. The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci. 2006;9:420–28. doi: 10.1038/nn1643.
34. Häfliger P. Adaptive wta with an analog vlsi neuromorphic learning chip. IEEE Trans Neural Netw. 2007;18(2):551–72.
35. Hahnloser R, Sarpeshkar R, Mahowald M, Douglas R, Seung S. Digital selection and analog amplification co-exist in an electronic circuit inspired by neocortex. Nature. 2000;405(6789):947–51.
36. Hansel D, Sompolinsky H. Methods in neuronal modeling, chap. Modeling feature selectivity in local cortical circuits. Cambridge, MA: MIT Press; 1998. p. 499–567.
37. Hylander P, Meador J, Frie E. VLSI implementation of pulse coded winner take all networks. In: Proceedings of the 36th Midwest symposium on circuits and systems, vol. 1. 1993. p. 758–61.
38. Indiveri G, Chicca E, Douglas R. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. IEEE Trans Neural Netw. 2006;17(1):211–21.
39. Indiveri G, Fusi S. Spike-based learning in VLSI networks of integrate-and-fire neurons. In: Proceedings on IEEE international symposium on circuits and systems, ISCAS 2007. 2007. p. 3371–4.
40. Indiveri G, Mürer R, Kramer J. Active vision using an analog VLSI model of selective attention. IEEE Trans Circuit Syst II. 2001;48(5):492–500.
41. Kandel ER, Schwartz J, Jessell TM. Principles of neural science. McGraw Hill; 2000.
42. Lazzaro J, Wawrzynek J, Mahowald M, Sivilotti M, Gillespie D. Silicon auditory processors as computer peripherals. IEEE Trans Neural Netw. 1993;4:523–8.
43. Lichtsteiner P, Posch C, Delbruck T. An 128 × 128 120 dB 15μs-latency temporal contrast vision sensor. IEEE J Solid State Circuit. 43(2):566–76.
44. Ma W, Beck J, Latham P, Pouget A. Bayesian inference with probabilistic population codes. Nat Neurosci. 2006;9(11):1432–8.
45. Mallik U, Vogelstein R, Culurciello E, Etienne-Cummings R, Cauwenberghs G. A real-time spike-domain sensory information processing system. In: Proceedings of IEEE international symposium on circuits and systems, vol. 3. 2005. p. 1919–22.
46. Markram H, Lübke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997;275:213–5.
47. Mead C. Analog VLSI and neural systems. Reading, MA: Addison-Wesley; 1989.
48. Mead C. Neuromorphic electronic systems. Proc IEEE. 1990;78(10):1629–36.
49. Merolla PA, Arthur JV, Shi BE, Boahen KA. Expandable networks for neuromorphic chips. IEEE Trans Circuit System I Fundam Theory Appl 2007;54(2):301–11.
50. Mitra S, Fusi S, Indiveri G. Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI. IEEE Trans Biomed Circuit Syst 2009;3(1), accepted Sept 2008, in press.

51. Oster M, Liu SC. A winner-take-all spiking network with spiking inputs. In: 11th IEEE international conference on electronics, circuits and systems (ICECS 2004). 2004.
52. Ott T, Stoop R. The neurodynamics of belief-propagation on binary markov random fields. In: Saul LK, Weiss Y, Bottou L, editors. Advances in neural information processing systems, vol. 18. Cambridge, MA: MIT Press; 2006. p. 1057–64.
53. Bofill-i Petit A, Murray A. Learning temporal correlations in biologically-inspired aVLSI. In: Proceedings of IEEE international symposium on circuits and systems, vol. V. IEEE; 2003. p. 817–20.
54. Bofill-i Petit A, Murray AF. Synchrony detection and amplification by silicon neurons with STDP synapses. IEEE Trans Neural Netw 2004;15(5):1296–304.
55. Rao P. Bayesian computation in recurrent neural circuits. Neural Comput. 2004;16:1–38.
56. Riis H, Hafliger P. Spike based learning with weak multi-level static memory. In: Proceedings of IEEE international symposium on circuits and systems. IEEE; 2004. p. 393–6.
57. Rutishauser U, Douglas R. State-dependent computation using coupled recurrent networks. Neural Comput. 2008; in press.
58. van Schaik A. Building blocks for electronic spiking neural networks. Neural Netw. 2001;14(6–7):617–28.
59. Schemmel J, Fieres J, Meier K. Wafer-scale integration of analog neural networks. In: Proceedings of the IEEE international joint conference on neural networks. 2008, in press.
60. Serrano-Gotarredona R, Oster M, Lichtsteiner P, Linares-Barranco A, Paz-Vicente R, Gómez-Rodríguez F, et al. AER building blocks for multi-layer multi-chip neuromorphic vision systems. In: Becker S, Thrun S, Obermayer K, editors. Advances in neural information processing systems, vol. 15. Cambridge, MA: MIT Press; 2005.
61. Somers DC, Nelson SB, Sur M. An emergent model of orientation selectivity in cat visual cortical simple cells. J Neurosci. 1995;15:5448–65.
62. Steimer A, Maass W, Douglas R. Belief-propagation in networks of spiking neurons. Neural Comput. 2009, submitted.
63. Systems of neuromorphic adaptive plastic scalable electronics (SyNAPSE). http://www.darpa.mil/dso/solicitations/baa08-28.htm. 2008.
64. Telluride neuromorphic cognition engineering workshop. http://ine-web.org/workshops/workshops-overview/.
65. van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science. 1996;274(5293):1724–6.
66. Wijekoon J, Dudek P. Compact silicon neuron circuit with spiking and bursting behaviour. Neural Netw. 2008;21(2–3):524–34.
67. Wilimzig C, Schneider S, Schöner G. The time course of saccadic decision making: dynamic field theory. Neural Netw. 2006;19:1059–74.
68. Yang Z, Murray A, Worgotter F, Cameron K, Boonsobhak V. A neuromorphic depth-from-motion vision model with stdp adaptation. IEEE Trans Neural Netw. 2006;17(2):482–95.
69. Yu A, Dayan P. Inference,attention, and decision in a Bayesian neural architecture. In: Saul LK, Weiss Y, Bottou L, editors. Advances in neural information processing systems, vol. 17. Cambridge, MA: MIT Press; 2005. p. 1577–84.