Workshop on Brain-Inspired Computing 2023

Modena

Workshop on Brain-Inspired Computing 2023, Modena

ISBN
9782832512340

DOI
10.3389/978-2-8325-1234-0

Citation
University of Modena and Reggio Emilia. (2023)

June 8–9, 2023, Modena

Table of contents

5 Welcome to the Workshop on Brain-Inspired Computing 2023

7 Insertion, biocompatibility and functionality of flexible neural implantable probes in *vivo*
Stefania Bartoletti, Elisa Ren, Federica Raimondi, Eve McGlynn, Maria Cerezo-Sanchez, Beatrice Casadei Garofani, Arianna Capodiferro, Hadi Heidari, Giulia Curia

12 CMOS leaky integrate-and-fire neuron circuit with memristor-based synapses reveals biologically plausible information transmission
Lorenzo Benatti, Tommaso Zanotti, Daniela Gandolfi, Jonathan Mapelli, Francesco Maria Puglisi

19 Full-scale point-neuron model of the mouse hippocampal microcircuits
Giulia Maria Boiani, Sergio Solinas, Giacomo Preti, Auror Manini, Michele Migliore, Jonathan Mapelli, Daniela Gandolfi

25 In-materia adaptive computing devices based on random-assembled clusters network
Francesca Borghi, Giacomo Nadalini, Silvia Bressan, Stefano Radice, Paolo Milani

29 Reservoir computing with a memristor tunable chaotic circuit
Manuel Escudero, Sabina Spiga, Stefano Brivio

33 Neuromorphic human activity recognition through LIF-based neurons
Vittorio Fra, Enrico Macii, Gianvito Urgese

38 Mesoscale "digital twins" for motor plan cancellation in nonhuman primates
Andrea Galluzzi, Nicola Alboré, Pierpaolo Pani, Stefano Ferraina, Maurizio Mattia

45 Bio-physics-based finite elements numerical simulation of an ionic emission and electrical sensing platform for neuron activity modulation: The case of calcium ions and calcium ion channels
Jacopo Nicolini, Pierpaolo Palestri, Luca Selmi

51 Constraint satisfaction problems solution through spiking neural networks with improved reliability: The case of Sudoku puzzles
Riccardo Pignari, Vittorio Fra, Evelina Forno, Enrico Macii, Gianvito Urgese

55 Resilience of an inferential neuromorphic circuit
Mirco Tincani, Giulia Maria Boiani, Lorenzo Benatti, Tommaso Zanotti, Francesco Maria Puglisi, Giuseppe Pagnoni, Jonathan Mapelli, Daniela Gandolfi

60 RRAM-based Binarized neural networks inference accelerator exploiting multiple in-memory computing paradigms
Tommaso Zanotti, Paolo Pavan, Francesco Maria Puglisi

Welcome to the Workshop on Brain-Inspired Computing 2023

One of the main challenges for the expansion of Artificial Intelligence is to draw inspiration from the way human brain works to develop new hardware emulating computation performed by neurons and synapses.

The workshop explores different aspects of neuromorphic computing: From biologically realistic models derived from experiments to synaptic and neuromorphic electronics to drive the latest generation of artificial neural networks.

This workshop will address these topics in a highly interdisciplinary scenario exploring areas of research apparently distant such as computational neuroscience and electronics, maintaining a common thread given by the performances shown by the nervous system, from energy consumption to highly parallelized cognitive tasks.

LIST OF ORGANIZERS

University of Modena and Reggio Emilia

Insertion, biocompatibility and functionality of flexible neural implantable probes in *vivo*

Author
Stefania Bartoletti – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
Elisa Ren – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
Federica Raimondi – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
Eve McGlynn – Microelectronics Lab, James Watt School of Engineering, Glasgow University, G12 8QQ, United Kingdom
Maria Cerezo-Sanchez – Microelectronics Lab, James Watt School of Engineering, Glasgow University, G12 8QQ, United Kingdom
Beatrice Casadei Garofani – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
Arianna Capodiferro – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
Hadi Heidari – Microelectronics Lab, James Watt School of Engineering, Glasgow University, G12 8QQ, United Kingdom
Giulia Curia – Laboratory of Experimental Electroencephalography and Neurophysiology, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy

Citation
Bartoletti, S., Ren, E., Raimondi, F., McGlynn, E., Cerezo-Sanchez, M., Garofani, B.C., Capodiferro, A., Heidari, H., Curia, G. Insertion, biocompatibility and functionality of flexible neural implantable probes in *vivo*.

Introduction

Neurostimulation technologies have been investigated as emerging therapeutic approaches for treating several neurological diseases, such as epilepsy, Parkinson's disease, and motor dysfunctions [1-2]. Although significant advancements have been made in the long-term stability of implants, challenges persist regarding their size, rigidity, and the methods used for their insertion into neural tissue [3]. Rigid stimulators and silicon-based devices are extensively used, but despite their beneficial qualities, their stiffness can cause inflammation as a response to mechanical stress [4]. To overcome these issues, probes have been developed with flexible substrates that reduce the impact on the brain tissue, but they introduced new challenges for surgical implant strategies [5].

To this aim, we have tested two implantation strategies to insert a custom-made flexible neuromorphic computing system (NCS) into the rat brain: *i*) probes coated with temporary stiffeners and *ii*) probes uncoated using a "sewing hole" strategy. Moreover, the biocompatibility of both coated and bare probes has been screened *in vivo*. In addition, NCS functionality has been tested *in vitro* and *in vivo* for stimulation and recording.

Methods

Probes are made up of three sheets of biopolymers: polyimide PI-2545, polyimide HD-4110, and a second layer of PI-2545. For *in vivo* tests, adult male Sprague Dawley rats underwent stereotaxic surgery and were implanted bilaterally in the deep brain areas with polymer-based NCS probes.

Two implantation strategies were tested: *i*) the NCS coated with bioresorbable stiffeners (5% silk fibroin, 1% alginate in 4 mM $CaCl_2$) and *ii*) a “sewing-hole” strategy of uncoated probes. The correct location of the implanted probe has been verified after 1 week of implantation by hematoxylin and eosin staining.

Unwanted immune reactions due to coated and uncoated probes were evaluated by immunofluorescence staining in the neocortex 3 weeks post-implantation. Positivity to markers for astrocytosis (GFAP) and microglia activity (Iba1) was measured to evaluate neuroinflammation, while neuronal loss was assessed quantifying NeuN-positive cells. Quantifications of GFAP and Iba-1 expression and $NeuN^{+}$ cell density (cell count per tissue area) were performed radially from the lesion in the range of 25 μm to 150 μm away from the lesion border zone and normalized using data from the concentric rings 450 μm away from the injection site [6].

Probe functionality was evaluated *in vitro* and *in vivo* measuring the signal in response to different stimulation protocols: stimulus width (100 μs in the monophasic setting; 100 μs with a 100 μs of zero time in the biphasic setting), amplitude (100 and 500 μA in vivo; 50, 100, 500 and 800 μA in vitro), frequency (1 and 130 Hz), train duration (1 min), with a 30-sec interval between each stimulation protocol.

Results

The “sewing hole” emerged as the best strategy to implant uncoated flexible probes more precisely into deep areas of the brain, while bioresorbable stiffeners did not provide sufficient probe rigidity for the reliable insertion into the rat brain.

Both microglia and astrocytes contribute to the immune response known as the foreign body reaction because of brain damage due to the insertion of a foreign substance. GFAP and Iba-1 expressions were higher around the lesion core and gradually decreased moving away from the electrode position for all coated and uncoated probes, as assessed by immunofluorescence. Astrocytic glial scar thickness did not exhibit differences among the different probe conditions. None of the devices adversely affected the distributions of neuronal cells all around the implanted probes.

The functional probe test has been performed *in vitro* and subsequently used *in vivo,* to confirm the capability of the NCS to stimulate and record within the desired stimulus protocols.

Conclusions

In this study, we identified a proper surgical procedure for the precise insertion into deep brain areas of an ultra-flexible functional neural device composed of polyimide layers of different thicknesses; in parallel, promising data collected from biocompatibility study and probe test performance suggested that our proposed probe presents noteworthy features as a potential candidate for long-term in vivo applications.

References

[1] Edwards, C.A., Kouzani, A., Lee, K.H., Ross, E.K., 2017. Neurostimulation Devices for the Treatment of Neurologic Disorders. Mayo Clinic Proceedings 92, 1427–1444. https://doi.org/10.1016/j.mayocp.2017.05.005

[2] Krauss, J.K., Lipsman, N., Aziz, T., Boutet, A., Brown, P., Chang, J.W., Davidson, B., Grill, W.M., Hariz, M.I., Horn, A., Schulder, M., Mammis, A., Tass, P.A., Volkmann, J., Lozano, A.M., 2021. Technology of deep brain stimulation: current status and future directions. Nat Rev Neurol 17, 75–87. https://doi.org/10.1038/s41582-020-00426-z

[3] I. Vėbraitė and Y. Hanein, 'Soft Devices for High-Resolution Neuro-Stimulation: The Interplay Between Low-Rigidity and Resolution', Front. Med. Technol., vol. 3, p. 675744, Jun. 2021, doi: 10.3389/fmedt.2021.675744.

[4] Gulino, M., Kim, D., Pané, S., Santos, S.D., Pêgo, A.P., 2019. Tissue Response to Neural Implants: The Use of Model Systems Toward New Design Solutions of Implantable Microelectrodes. Front. Neurosci. 13, 689. https://doi.org/10.3389/fnins.2019.00689

[5] Luo, J., Xue, N., Chen, J., 2022. A Review: Research Progress of Neural Probes for Brain Research and Brain–Computer Interface. Biosensors 12, 1167. https://doi.org/10.3390/bios12121167

[6] Wellman SM, Li L, Yaxiaer Y, McNamara I, Kozai TDY. Revealing Spatial and Temporal Patterns of Cell Death, Glial Proliferation, and Blood-Brain Barrier Dysfunction Around Implanted Intracortical Neural Interfaces. Front Neurosci. 13:493. doi: 10.3389/fnins.2019.00493. May. 2019.

CMOS leaky integrate-and-fire neuron circuit with memristor-based synapses reveals biologically plausible information transmission

Author
Lorenzo Benatti – Dipartimento di Ingegneria "Enzo Ferrari", Via. P. Vivarelli 10/1, 41125 Modena, Italy, Università degli Studi di Modena e Reggio Emilia
Tommaso Zanotti – Dipartimento di Ingegneria "Enzo Ferrari", Via. P. Vivarelli 10/1, 41125 Modena, Italy, Università degli Studi di Modena e Reggio Emilia
Daniela Gandolfi – Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze, Università degli Studi di Modena e Reggio Emilia; Centro Interdipartimentale di Neuroscienze e Neurotecnologie, Università degli Studi di Modena e Reggio Emilia
Jonathan Mapelli – Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze, Università degli Studi di Modena e Reggio Emilia; Centro Interdipartimentale di Neuroscienze e Neurotecnologie, Università degli Studi di Modena e Reggio Emilia
Francesco Maria Puglisi – Dipartimento di Ingegneria "Enzo Ferrari", Via. P. Vivarelli 10/1, 41125 Modena, Italy, Università degli Studi di Modena e Reggio Emilia; Centro Interdipartimentale di Neuroscienze e Neurotecnologie, Università degli Studi di Modena e Reggio Emilia

Citation
Benatti, L., Zanotti, T., Gandolfi, D., Mapelli, J., Puglisi, F.M. CMOS leaky integrate-and-fire neuron circuit with memristor-based synapses reveals biologically plausible information transmission.

Introduction

Spiking neuromorphic circuits mimic energy-efficient mechanisms of biological systems, offering potential for brain-like computation with low power consumption in electronic implementations. Spiking Neural Networks (SNNs) transmit binary spikes through synapses that undergo persistent changes in strength based on stimulation patterns. This synaptic plasticity [1] can be reliably emulated by memristors [2], thanks to their ability to adjust their conductance according to past activity. These advancements enabled efficient and biologically inspired implementation of artificial neural networks. However, it is uncertain if current neuromorphic hardware can effectively transfer information like biological networks. In neuronal microcircuits, information is exchanged through input and output spike series. Shannon's Mutual Information (MI) estimates the information transmitted by circuits, neurons, or synapses without requiring knowledge of the neural code. MI is calculated using response and noise entropy, reflecting response variability. Subtracting noise entropy from response entropy measures the system's information transmission capacity. MI is measured in bits and is computed using the following equation:

$$MI(R,S) = MI(S,R) = H(R) - H(R|S) = \sum_{s \in S} \sum_{r \in R} p(s) p(r|s) \log_2 \frac{p(r|s)}{p(r)} \tag{1}$$

where *r* and *s* are the response and the stimulus pattern, respectively; *p(r)* and *p(s)* are the probabilities that *r* or *s* occur simultaneously within the same time window (hereafter called bin and expressed in *ms*). Finally, *p(r|s)* is the probability of obtaining the response pattern *r* given the stimulus pattern *s*.

The MI can be decomposed in the single stimulus contribution (Stimulus Specific Surprise - SSS) or even single spike contribution (Surprise per Spike – SpS):

$$I(s^*) = SSS = \sum_{r} p(r|s^*) \log_2 \frac{p(r|s^*)}{p(r)} \tag{2}$$

$$I_{perSpike}(s^*) = SpS = \frac{I(s^*)}{stimuls\ spike\ count} \tag{3}$$

Synaptic devices

We investigated information propagation in a CMOS-based artificial cerebellar granule cell with four memristor-based synaptic inputs. Through circuit simulations using Cadence Virtuoso software, we accurately modelled the artificial CMOS neuron's response using Verilog-A behavioral descriptions. The memristive elements (commercially available C-doped Self-Directed Channel (SDC) memristors from Knowm [3]) were modelled using the UniMORE RRAM compact model [4], which incorporates physics-based principles and is corroborated by the results of advanced multiscale simulations [5]. Our findings demonstrate that MI propagates in a CMOS Leaky Integrate and Fire (LIF) [6] neuron-based SNN with memristive synapses in a way that resembles the one found in biological networks.

Neuron model and simulated network

In the LIF neuron model (Fig. 1a), synaptic memristive devices exhibit rate-dependent plasticity. Higher (lower) presynaptic stimulation rates lead to potentiation (depression) of corresponding synapses (Fig. 1b). Input spikes from presynaptic neurons are integrated into a capacitor with a virtual ground terminal. When the capacitor's voltage surpasses a threshold, an output spike is generated. After a predetermined delay (T_{spike} Delay), the capacitor is discharged to reset the system. The charging rate of the capacitor depends on the input spike rate and presynaptic strength. In the absence of input spikes, the neuron discharges the capacitor with a specific time constant, mimicking the depolarization caused by the leaky membrane in biological neurons. In our neuron model, this learning rule is implemented by shaping the spike waveform (Fig. 1c), resulting in small potentiation or depression of associated synaptic memristor devices. To evaluate information transmission similarities between artificial SNNs and biological systems, we constructed a circuit resembling granule cell morphology and implemented spike digitization principles (Fig. 1d). Spike stimulations lasting 10ms were randomly applied within 50ms time bins, capturing the stochastic nature observed in biological networks within an inherently deterministic neuron. This study aimed at quantifying the impact of learning on theoretical quantities related to information propagation. In biological

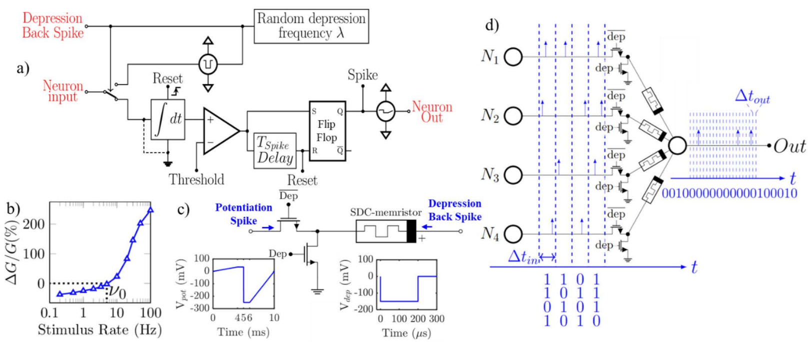


FIGURE 1

(a) LIF neuron model. **(b)** Synaptic plasticity implemented through a rate-based learning rule: higher (lower) presynaptic stimulation rate than ν_0 leads to potentiation (depression) of the synapse. **(c)** Spiking diagram illustrating potentiation and depression. **(d)** Implementation of a biologically plausible granule cell network with 4 input neurons (N_{1-4}) and an output neuron using memristor-based synapses. Input time is divided into Δt_{in}=50ms time bins with stochastic spike distribution, and output time with Δt_{out}=10ms.

experiments, synaptic efficacy is measured as release probability ($\overline{p}$) [7], while memristor-based neuromorphic networks use memristor conductance as weight representation. We simulated the system with varying memristor conductance to analyze the effects on mutual MI, SSS, and SpS due to synaptic plasticity. To focus only on information propagation in an artificial network, we fixed synaptic strength and disabled plasticity during input spike train application, enabling it only when delivering theta bursts (Fig. 2a-b) to change the synaptic strength. Thus, the information transfer analysis results outlined in the following section remain valid, regardless of the specific device used to represent the synaptic weight.

Results and Discussion

We can examine the impact of synaptic plasticity on a spiking neuron with memristive synapses by analyzing key information transfer quantities like entropy, MI, and surprise. Following the approach in [7, 8], we digitized

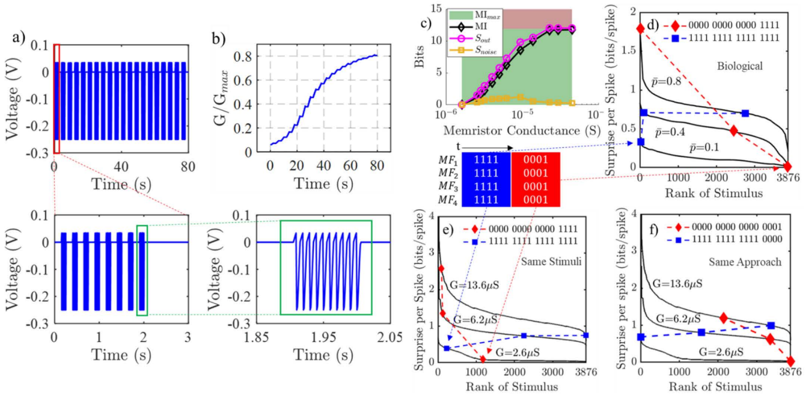


FIGURE 2

(a) Plasticity simulation involves a potentiation theta burst train, resulting in a conductivity change. **(b)** The conductivity change is used to adjust the synaptic weights of the network. In **(a)**, the LTP theta burst sequence is shown at different zoom levels. In **(b)**, the normalized weight change over time is depicted when delivering the pattern from **(a)**. **(c)** Mutual information (black), output entropy (S_{out}, magenta), and noise entropy (S_{noise}, yellow) versus memristor conductance. **(d)** The surprise per spike (SpS) is plotted against the release probability ($\overline{p}$), adapted from [4]. The stimuli are ranked based on their SpS for each $\overline{p}$ value of 0.1, 0.4, and 0.8. Blue and red stimuli, representing the maximum and minimum SpS for the most depressed network ($\overline{p}$ = 0.1), are monitored for different release probabilities. The same experiment is conducted in our network with varying synapse conductance values, tracking the same stimuli as in **(d).** The results in **(e)** and **(f)** exhibit similar trends as in **(d)**.

spike trains and selected a controlled set of stimuli 'S'. We then detected responses 'r' when presenting stimuli with known probabilities *p(s)*. By collecting the data, we estimated the joint probabilities *p(r|s)* and the average probability distribution of responses *p(r)* across stimuli. We then computed MI, which is known to vary with release probability in biological systems [7]. We investigated the relationship between MI and memristor conductance, representing synaptic efficacy. Fig. 2c illustrates the correlation between MI and memristor conductance values. Increasing synaptic efficacy (memristor conductance) enhances overall information transfer, aligning with expectations [7]. We also aimed at identifying the most effectively encoded stimuli by the electronic neuron. Stimulus-specific contribution to MI (SSS; eq. 2) and SpS (eq. 3) were computed to identify the informative set of stimuli. Following [7] (Fig. 2d), for each synaptic strength value, the network was stimulated with 3876 inputs, ranked by descending SpS (black curves in Fig. 2e-f). This procedure was repeated for different memristor conductance values (2.6 µS, 6.2 µS, 13.6 µS) as shown in Fig. 2e-f. Initially, we focused on the results obtained with the lowest memristor conductance value. The stimuli with the highest and lowest SpS were identified (blue and red markers, respectively). When synaptic potentiation was applied, these markers moved in line with previous findings [7], as shown in Fig. 2d. Although the specific stimuli did not match those in [7] at the lowest $\overline{p}$ value, the qualitative trend was consistent. These findings demonstrate the reliability of the artificial memristor-based neuro-synaptic circuit in quantifying the information content of a specific spike train based on network strength.

Conclusions

In this study, we examined the similarities between an artificial neuron with memristor synapses and a rate-based learning rule, and the response of a biological neuron in terms of information propagation. By linking the biological release probability $\overline{p}$ with the conductance of artificial synapses, we computed MI, SSS, and SpS for different synaptic weights. The results demonstrate that the implemented neuron can effectively retrieve and quantify information from a specific stimulus based on synaptic strength, resembling biological networks, while offering advantages in terms of time scale, area, and applicability in various application domains.

References

[1] Mapelli J, Gandolfi D, Vilella A, Zoli M, Bigiani A. Heterosynaptic GABAergic plasticity bidirectionally driven by the activity of pre- and postsynaptic NMDA receptors. Proc Natl Acad Sci U S A. 2016 Aug 30;113(35):9898-903. doi: 10.1073/pnas.1601194113. Epub 2016 Aug 16. PMID: 27531957; PMCID: PMC5024594.

[2] Mario Lanza et al. “Memristive technologies for data storage, computation, encryption, and radio-frequency communication” Science 376, eabj9979(2022). DOI:10.1126/science.abj9979.

[3] K.A. Campbell, “Self-directed channel memristor for high temperature operation,” in Microelectronics Journal, vol. 59, Jan. 2017, pp. 10-14, doi: 10.1016/j.mejo.2016.11.006.

[4] F.M. Puglisi, T. Zanotti and P. Pavan, “Unimore Resistive Random Access Memory (RRAM) Verilog-A Model,” nanoHUB, Jun 2019, doi: 10.21981/15GF-KX29.

[5] A. Padovani, L. Larcher, F. M. Puglisi and P. Pavan, “Multiscale modeling of defect-related phenomena in high-k based logic and memory devices,” in 2017 IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), Jul. 2017, pp. 1–6, doi: 10.1109/IPFA.2017.8060063

[6] D. Florini, D. Gandolfi, J. Mapelli, L. Benatti, P. Pavan and F. M. Puglisi, “A Hybrid CMOS-Memristor Spiking Neural Network Supporting Multiple Learning Rules,” in IEEE Transactions on Neural Networks and Learning Systems, 2022, doi: 10.1109/TNNLS.2022.3202501.

[7] Arleo A, Nieus T, Bezzi M, D’Errico A, D’Angelo E, Coenen OJ. How synaptic release probability shapes neuronal transmission: information-theoretic analysis in a cerebellar granule cell. Neural Comput. 2010; 22(8): 2031–58. pmid:20438336

[8] Mapelli J, Gandolfi D, Giuliani E, Prencipe FP, Pellati F, Barbieri A, D’Angelo E, Bigiani A. The effect of desflurane on neuronal communication at a central synapse. PLoS One. 2015 Apr 7;10(4):e0123534. doi: 10.1371/journal.pone.0123534. PMID: 25849222; PMCID: PMC4388506.

Full-scale point-neuron model of the mouse hippocampal microcircuits

Author
Giulia Maria Boiani – Department of Biomedical, Metabolic and Neural Sciences, University of Modena, Italy
Sergio Solinas – Department of Biomedical Sciences, University of Sassari, Sassari, Italy
Giacomo Preti – Department of Biomedical, Metabolic and Neural Sciences, University of Modena, Italy
Auror Manini – Department of Biomedical, Metabolic and Neural Sciences, University of Modena, Italy
Michele Migliore – Institute of Biophysics, National Research Council, Palermo, Italy
Jonathan Mapelli – Department of Biomedical, Metabolic and Neural Sciences, University of Modena, Italy; Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Italy
Daniela Gandolfi – Department of Biomedical, Metabolic and Neural Sciences, University of Modena, Italy; Center for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Italy

Citation
Boiani, G.M., Solinas, S., Preti, G., Manini, A., Migliore, M., Mapelli, J., Gandolfi, D. Full-scale point-neuron model of the mouse hippocampal microcircuits.

Introduction

Spiking neural networks are the best candidates for the creation of full-scale point networks of brain areas. These models are becoming a reliable tool for investigating brain activity and study dysfunctions as well as for building digital twins of brain areas under investigation [1].

One of the main critical points of point neurons models is the computational cost associated with the computation of the connectome, since in many brain areas connectivity is shaped by morpho-anatomical constraints and is not properly described by isotropic random connection strategies [2].

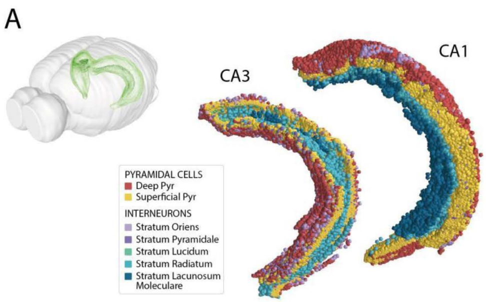
We proposed a connection strategy based on the description of neurites through probability density functions, which respect the citoarchitecture. This strategy was benchmarked in the CA1 area of the mouse hyppocampus [3] and applied for modelling the CA3 area, but can be reliably extended to other brain areas and translated to model human brain networks [4]. Here we describe the modelling pipeline for the creation of a full-scale single cell resolution model of hippocampal microcircuits.

Results

The modelling procedure was data-driven, meaning that placement data and morphology information about neurons were taken from literature and cell atlases. The pipeline consisted of the following steps:

A) Placement: data of cell positioning were taken from the Blue Brain Cell Atlas database (https://bbp.epfl.ch/nexus/cellhttps://bbp.epfl.ch/nexus/cell-atlas/atlas/). The atlas provides placement data of excitatory and inhibitory neurons distributed in the layers of CA1 and CA3 areas. Inhibitory neurons were further divided according to immunolabelling expression. We combined these data with information obtained from literature allowing the identification of 11 and 12 interneurons types for CA1 and CA3 area respectively. Moreover, 2 pyramidal cell types were generated and placed within layers according to soma positioning, see Fig.1 [5,6].

B) Morphology description: axons and dendrites were described through probability density clouds with combinations of geometrical shapes (cones and ellipsoids), see Fig 2. Pyramidal cells of the CA3 have 2 distinct axonal arborizations, spanning in the stratum oriens and stratum radiatum. The latter was described with a tubular shape to better fit anatomical constraints, see Fig.2B. Average dimensions of neurites were calculated



B

NEURONAL POPULATION

Excitatory neurons

	CA1		CA3	
Stratum	Cell type	n	Cell type	n
SO	Deep Pyr	45408	Deep Pyr	38512
SP	Superficial Pyr	216435	Superficial Pyr	359885
	Tot	261843	Tot	398397

Inhibitory neurons

	CA1		CA3	
Stratum	Cell type	n	Cell type	n
SO	OLM, TRI, BP	1979	OLM, TRI	4076
SP	CCKBC, PVBC, AA	1661	CCKBC, PVBC, AA, BS, IVY	5254
SL			SL, MFA	693
SR	IVY, BS, SCA, PPA	10147	R, PPA	3718
SLM	NGF	12397	NGF	168
	Tot	26184	Tot	13909

FIGURE 1
(A) *left* The mouse hippocampus is shown in green. *Right* Plot of the scaffolds of the CA1 and CA3 areas, each point representing a neuron. Cells are divided according to the layer in which their soma is located.
(B) Neuronal count for each layer of the CA1 and CA3 areas. (SO: stratum oriens; SP: stratum pyramidale; SL: stratum lucidum; SR: stratum radiatum; SLM: stratum lacunosum moleculare. Pyr, Pyramidal cell; OLM, Oriens-Lacunosum Molecolare cell; TRI, Trilaminar cell; BP, Bipolar cell; CCKBC, CCK+ Basket cell; PVBC, PV+ Basket cell; AA, Axo-axonic cell; BS, Bistratified cell; IVY, Ivy cell; SCA, Schaffer's collateral-associated cell; PPA, Perforant path-associated cell; SL, Spiny lucidum cell; MFA, Mossy fibers-associated cell; R, Radiatum cell; NGF, Neurogliaform cell.).

according to literature and morphology atlases like neuromorpho.org (https://neuromorpho.org/) or Janelia Mouselight (http://ml-neuron-browser.janelia.org/). Furthermore, cell-to-cell variability was included. Neurons were then placed automatically in the 3D space orienting the neurites according to external landmarks.

C) Connectivity: Axons were then converted into convex hulls and the intersections between axons and dendritic point clouds of each neuron were computed. The putative number of connecting neurons was preliminarily assessed by calculating the intersections of bounding boxes. The resultant connection pairs were then pruned according to literature synaptic connection probabilities.

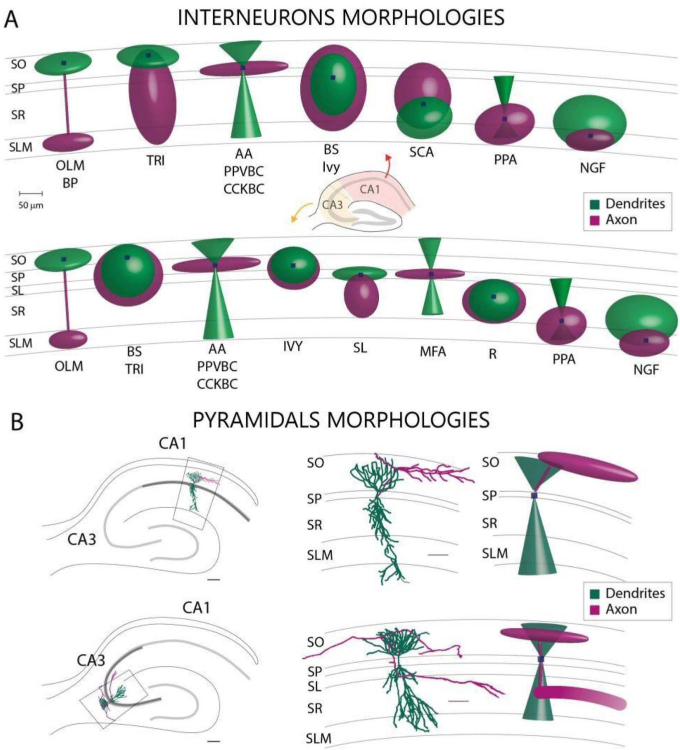


FIGURE 2

(A) Probability density functions of GABAergic interneurons. The 11 classes of CA1 interneurons described in table 1 have been represented with 7 shapes. The 12 classes of CA3 interneurons described in Table 1 have been represented with 9 shapes. Black squares indicate the preferential position of the cell soma while green and violet shades represent dendrites and axons respectively. **(B)** Pyramidal neurons. *left* Scheme of a transversal section of the mouse hippocampus. Thick gray lines represent Pyramidal layers of both CA1 and CA3. *right* Realistic morphology of a PC (dendrites in green, and axon in violet), placed as seen in inset on the left, whith their corresponding probability density functions, respecting the color code. Scale bars in µm (100 left, 50 right).

D) Simulation: the full-scale network was simulated with NEST, a point neuron simulator, selecting and tuning proper neuron and synapse models. Here an integrate-and fire neuron model with adaptive threshold, namely a Hill-Tononi model, was employed and neurons were endowed with short term synaptic plasticity, using a Tsodyks-Markram model for the synapse [7,8].

Conclusions

This modelling pipeline is a generalized bottom-up strategy to build networks of point neurons based on morpho-anatomical constraints. It was employed for the creation of a full-scale point neuron model of the mouse CA1 hippocampus [3] and then adopted to generate the CA3 area. The method is based on the construction of geometrical probability density functions mimicking axonal and dendritic shapes and extensions. It can be helpful for modelling microcircuits where data on cell morphologies are poor or absent and can be easily translated to other brain areas. Therefore, it is likely to be a promising tool for the construction of large-scale circuits with realistic connectome that can be used to investigate physiological and pathological conditions (e.g temporal lobe epilepsy [9]) at single neuron level.

References

[1] Markram H et al. Reconstruction and Simulation of Neocortical Microcircuitry. Cell. 2015

[2] Giacopelli G et al. On the structural connectivity of large-scale models of brain networks at cellular level Sci Rep. 2021

[3] Gandolfi D et al. Full-scale scaffold model of the human hippocampus CA1 area. Nat Comput Sci, 2023.

[4] Gandolfi D et al. A realistic morpho anatomical connection strategy for modelling full-scale point-neuron microcircuits. Sci Rep. 2022

[5] Le Duigou C et al. Recurrent synapses and circuits in the CA3 region of the hippocampus: an associative network. Front Cell Neurosci, 2014

[6] Booker SA et al. Morphological diversity and connectivity of hippocampal interneurons. Cell and Tissue Research, 2018

[7] Hill S, Tononi G Modeling sleep and wakefulness in the thalamocortical system. Journal of Neurophysiology. 2005

[8] Tsodyks M et al. Synchrony generation in recurrent networks with frequency-dependent synapses. Journal of Neuroscience 2000

[9] Wang HE et al. Delineating epileptogenic networks using brain imaging data and personalized modelling in drugresistant epilepsy. Sci Transl Med. 2023

In-materia adaptive computing devices based on random-assembled clusters network

Author
Francesca Borghi, Giacomo Nadalini, Silvia Bressan, Stefano Radice, Paolo Milani – CIMAINA and Dipartimento di Fisica, Università di Milano, Italy

Citation
Borghi, F., Nadalini, G., Bressan, S., Radice, S., Milani, P. In-materia adaptive computing devices based on random-assembled clusters network.

Introduction

Artificial neuromorphic systems aim at the reproduction of the brain performances in terms of classification and pattern recognition, typically using components obtained by top-down lithographic technologies regularly employed by digital computers.[1] Unconventional in-materia computing has been proposed as an alternative strategy, exploiting the complexity and collective phenomena originating from various classes of physical substrates.[2,3]

The employment of materials composed by a large number of non-linear nanoscale junctions, random assembled according to a non-deterministic strategy on a substrate, are of particular interest for the implementation of in-materia computing systems.[4,5] In particular, the employment of random-assembled memristive materials is a strategic solution for the development of energy efficient and neuromorphic computing devices, well beyond the von Neuman bottleneck.[6]

In this contest, films obtained by the assembling of metallic nanoparticles and, in particular, Au cluster-assembled films have shown interesting non-linear electrical properties and complex resistive switching phenomena.[7-9] This has been exploited for the fabrication of devices (Receptrons) able to perform binary classification of boolean functions.[10]

However, in loco processing of analog data and the need of computing unit integrated with sensor devices in the same compact hardware system exceed the Boolean algebra and the information coding based on binary inputs. Furthermore, hardware computing solutions able to perceive and to adapt to external stimuli are strategic for the implementation of embodied intelligent system.

Results

Here we demonstrate the development of a memristive switching device based on nanostructured gold able to integrate and process voltage pulse trains autonomously. In particular, the device performs durable, controlled and reversible resistive switching events between two main resistance levels, separated by four orders of magnitude, holding after hundreds of cycles. The reversible change in resistance is related to the intensity and the temporal features of the voltage pulses. Interestingly, the two resistance levels correspond to different conduction mechanisms, from ohmic to a non-linear one. Furthermore, we prove the integration of such a type of memristive device into a polymer matrix of polydimethylsiloxane (PDMS), which is one of the most common flexible and stretchable silicon elastomers, widely used for stretchable electronics and biomedical applications, microfluidic devices, prostheses, as also as food additive. The composite Au/PDMS device preserves the electrical conductivity even in stress conditions and adapts it in response to the different mechanical stimuli the system receives in a non-trivial manner. Even more interestingly, the memristive flexible and stretchable composite device processes and integrates differently specific train pulses according to the external mechanical stimuli applied to the device.

Conclusions

We here propose the integration of a resistive switching nanostructured metallic film into a flexible and stretchable medium, as polydimethylsiloxane (PDMS). The computation capability and the integration of voltage pulse trains is here reported for the memristive composite device, as the controlled and reversible adaptation of the computational capability of the system according to the external mechanical stimuli received. The employment of such a neuromorphic device is proposed as a representative key solution for the development of embodied intelligent system.

References

[1] Zheng, N.; Mazumder, P. *Learning in Energy-Efficient Neuromorphic Computing : Algorithm and Architecture Co-Design*; Wiley, 2020.

[2] Stepney, S.; Rasmussen, S.; Amos, M. *Computational Matter*, 1st ed.; Springer Cham, 2018. https://doi.org/10.1007/978-3-319-65826-1.

[3] Jaeger, H. Towards a Generalized Theory Comprising Digital, Neuromorphic and Unconventional Computing. *Neuromorphic Comput. Eng.* **2021**, *1* (1), 012002. https://doi.org/10.1088/2634-4386/abf151.

[4] Sillin, H. O.; Aguilera, R.; Shieh, H. H.; Avizienis, A. V.; Aono, M.; Stieg, A. Z.; Gimzewski, J. K. A Theoretical and Experimental Study of Neuromorphic Atomic Switch Networks for Reservoir Computing. *Nanotechnology* **2013**, *24* (38). https://doi.org/10.1088/0957-4484/24/38/384004.

[5] Milano, G.; Montano, K.; Ricciardi, C. In Materia Implementation Strategies of Physical Reservoir Computing with Memristive Nanonetworks. *J. Phys. D. Appl. Phys.* **2023**, *56* (8). https://doi.org/10.1088/1361-6463/acb7ff.

[6] Bose, S. K.; Lawrence, C. P.; Liu, Z.; Makarenko, K. S.; J van Damme, R. M.; Broersma, H. J.; van der Wiel, W. G. Evolution of a Designless Nanoparticle Network into Reconfigurable Boolean Logic. *Nat. Nanotechnol.* | **2015**, *10*. https://doi.org/10.1038/NNANO.2015.207.

[7] Mirigliano, M.; Decastri, D.; Pullia, A.; Dellasega, D.; Casu, A.; Falqui, A.; Milani, P. Complex Electrical Spiking Activity in Resistive Switching Nanostructured Au Two-Terminal Devices. **2020**. https://doi.org/10.1088/1361-6528/ab76ec.

[8] Mirigliano, M.; Borghi, F.; Podestà, A.; Antidormi, A.; Colombo, L.; Milani, P. Non-Ohmic Behavior and Resistive Switching of Au Cluster-Assembled Films beyond the Percolation Threshold. *Nanoscale Adv.* **2019**, *1* (8). https://doi.org/10.1039/c9na00256a.

[9] Borghi, F.; Mirigliano, M.; Dellasega, D.; Milani, P. Influence of the Nanostructure on the Electric Transport Properties of Resistive Switching Cluster-Assembled Gold Films. *Appl. Surf. Sci.* **2022**, *582* (October 2021), 152485. https://doi.org/10.1016/j.apsusc.2022.152485.

[10] Mirigliano, M.; Paroli, B.; Martini, G.; Fedrizzi, M.; Falqui, A.; Casu, A.; Milani, P. A Binary Classifier Based on a Reconfigurable Dense Network of Metallic Nanojunctions. *Neuromorphic Comput. Eng.* **2021**, *1* (2), 024007. https://doi.org/10.1088/2634-4386/AC29C9.

Reservoir computing with a memristor tunable chaotic circuit

Author
Manuel Escudero, Sabina Spiga, Stefano Brivio – CNR – IMM, Unit of Agrate Brianza, 20864, ITALY

Citation
Escudero, M., Spiga, S., Brivio, S. Reservoir computing with a memristor tunable chaotic circuit.

Introduction

Reservoir computing is a computational framework derived from recurrent neural networks, and also inspired by the activity in human brain [1]. It consists in a reservoir that maps the input to a high dimensional space, and a readout layer that linearly transforms the reservoir output. The main advantage of reservoir computing is that only the weights of the readout layer need to be trained, reducing the training cost while expanding the alternatives for implementing the reservoir. For this reason, physical reservoir computing has been gaining considerable attention [2], with a special interest in further minimizing implementation costs and build extremely compact reservoirs [3]-[5]. It is believed that reservoirs typically perform best when they work in the 'edge of the chaos', i.e., in the transition between ordered and disordered dynamics. However, there is evidence that chaotic systems can also be considered for reservoir computing [6]. Based on these findings, we explore a memristor-based version of the Murali-Lakshamanan-Chua (MLC) circuit, as a compact reservoir that exhibits chaos and holds tunable properties.

Physical reservoir

The memristor-based MLC circuit (fig. 1) uses a nonvolatile $Pt/HfO_2/TiN$ memristor as a tunable nonlinear resistor to generate complex dynamics including chaos, fundamental for computation in a reservoir approach, while reducing the complexity of the original circuit [6]. More important, the nonvolatile memristor can be programmed to modify the dynamics of the circuit, granting to the circuit tunability features. Following similar design considerations as in [7], we designed and physically implemented the circuit in a breadboard. A symmetric square signal of variable amplitude F and fixed period T is applied as input, while the output oscillations are generated at the voltage across the capacitor v. Experimental results demonstrate that by increasing the forcing signal amplitude F, oscillations with periods multiple to T or even chaotic oscillations are generated (fig. 1).

Evaluation of tasks

The oscillator is employed as a tunable reservoir to evaluate two nonlinear tasks: logic functions and nonlinear classification. The complete scheme of the reservoir computing is reported in fig. 2, illustrating an example that

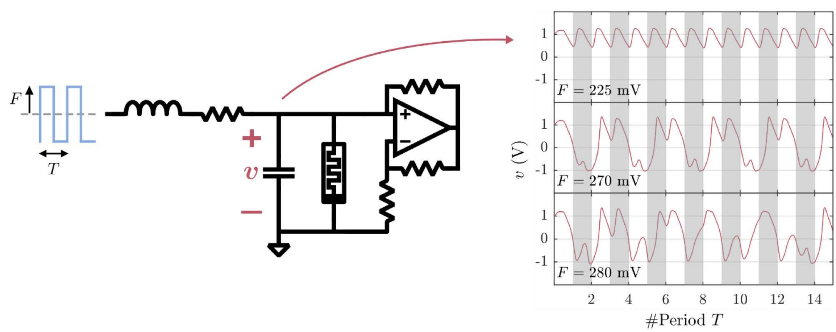


FIGURE 1
Memristor-based MLC circuit schematic driven with a square signal (left), and examples of experimental oscillations generated with the hardware implementation using different input signal amplitudes *F* (right).

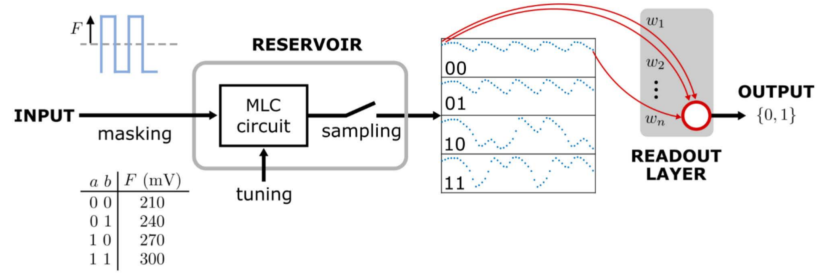


FIGURE 2

Reservoir computing scheme, illustrating the computation of a two-input logic function *f*(*a*, *b*). The input, a pair of Boolean values, is masked in the amplitude *F* of a square signal using a linear transformation (see table). The square signal is fed to the MLC circuit and the circuit output is sampled (an experimental result for each input is shown). The samples are then weighted sum in the readout layer, classifying the input as 0 or 1.

computes a two-input logic function. The input data are masked into the square signal amplitude *F* using a linear transformation. The resulting signal is applied to the circuit, and the oscillator output is sampled (dots in fig.2). The obtained sequence of samples is then weighted summed in the readout layer, thus obtaining the output. Between input evaluations, the circuit is reset to the same initial conditions, i.e., previous evaluations do not influence in future ones. We evaluated each task by generating ~4000 examples of circuit outputs used for the training (80%) and the validation (20%) of the readout layer. While the quantity of examples may seem redundant for tasks such logic functions, the presence of noise and the chaotic nature of the circuit may turn out in different reservoir outputs for the same task inputs. The experimental results comprised the acquisition of the MLC circuit sampled output for all the examples; for each input, 1000 samples were acquired over the course of 20 periods of the input signal (sampling rate = *T*/50). The readout layer training and validation of the task was performed off-line. Preliminary results reveal that the chaotic reservoir performs with a reasonable high accuracy (90-100%), even with the presence of noise. After analyzing the trained readout layer weights, we observed that most of the samples were either irrelevant or redundant in some of the evaluated tasks, indicating the possibility of aggressively scaling down the readout layer size.

Finally, we confirmed through simulations that the reservoir computing performance can be optimized by tuning our MLC circuit.

Conclusions

A memristor-based MLC circuit was physically implemented as a compact tunable reservoir with chaotic dynamics. Simulation and experimental results confirm the feasibility of the proposed memristor-based MLC circuit as a compact reservoir for computing small classification tasks, and the circuit tunable properties granted by the nonvolatile memristor can be exploited to improve the accuracy of the computation.

References

[1] P. Enel et al, PLoS Comput. Biol., vol. 12, n. 6, e1004967, 2016.

[2] G. Tanaka et al., Neural Networks, vol. 115, pp. 100–123, 2019.

[3] M.C.Soriano et al., IEEE Trans. Neural Networks Learning Systems, vol. 26, n. 2, pp. 388-393, 2015.

[4] D. Marković et al., Appl. Phys. Lett., vol. 114, 012409, 2019.

[5] M. R. E. U. Shougat et al., Sci. Rep., vol. 11, 19465, 2021.

[6] J.H. Jensen and G. Tufte, Proc. 14th European Conference on Artificial Life , pp. 222–229, 2017.

[7] M. Escudero et al., IEEE 48th European Solid-State Circuits Conference (ESSCIRC), p. 117-120, 2022.

Neuromorphic human activity recognition through LIF-based neurons

Author
Vittorio Fra, Enrico Macii, Gianvito Urgese – Politecnico di Torino, Electronic Design Automation (EDA) Group, Torino, Italy

Citation
Fra, V., Macii, E., Urgese, G. Neuromorphic human activity recognition through LIF-based neurons.

Introduction

Human Activity Recognition (HAR) is the classification task focused on daily-life activities. It relies on time-dependent signals coming from body monitoring, and it can be performed in different domains ranging from healthcare to sport and working conditions [1]. Consequently, HAR is inherently well suited for on-edge deployment of tailored services, whose effectiveness must be evaluated taking into account energy efficiency as an additional, major figure of merit. Bringing HAR into the neuromorphic domain can lead to significant improvements in this regard, especially if dedicated hardware is used [2]. An example of this latter is the Intel's Loihi chip [3,4], which implements a discrete Current-Based Leaky Integrate-and-Fire (CuBa-LIF) neuron model also referred to as second-order LIF model [5]. In view of on-edge applications of such neuromorphic hardware, it can be useful to investigate the suitability of a fully LIF-based Spiking Neural Network (SNN) in terms of both accuracy performance and energy consumption. Additionally, by working with such network, it is possible to investigate the possibility of targeting HAR by means of SNNs relying on the most biologically plausible neuron model among the less computational expensive ones [6].

Methods

As specific use case of the HAR task, we selected the subset of smartwatch signals for general, hand-oriented activities from the WISDM-HAR dataset [7] as in [1]. A three-layers, fully connected, recurrent SNN (RSNN) was used by employing two different LIF-based neuron models. Specifically, the input layer contains first-order LIF neurons, while recurrent second-order LIF neurons build up the hidden layer and the output layer consists of non-recurrent CuBa-LIF neurons. First-order and second-order models are described in time by eqn. (1), (2) and eqn. (3)-(5) respectively:

$$U_{t+1} = \beta \cdot U_t + W \cdot X_{t+1} - S_t \cdot \vartheta \tag{1}$$

$$\text{with } S_t = \begin{cases} 1, & if\ U_t > \vartheta \\ 0, & otherwise \end{cases} \tag{2}$$

$$U_{t+1} = \beta \cdot U_t + I_{syn,t+1} - S_t \cdot \vartheta \tag{3}$$

$$\text{with } I_{syn,t+1} = \alpha \cdot I_{syn,t} + W \cdot X_{t+1} \tag{4}$$

$$\text{and } S_t = \begin{cases} 1, & if\ U_t > \vartheta \\ 0, & otherwise \end{cases} \tag{5}$$

where t identifies the simulation timestep while $\alpha = e^{-1/\tau_{syn}}$ and $\beta = e^{-1/\tau_{mem}}$ are the decay rate for the synaptic current and for the membrane potential respectively, with the corresponding time constant τ_{syn} and τ_{mem}; I_{syn} is the synaptic current, W is the weights matrix and S_t represents spike emission, which occurs when the membrane potential U overcomes the threshold voltage ϑ.

Compared to [2], which used the same network architecture, we here did not perform signal encoding before feeding the RSNN. Raw signals were directly provided to the network, whose input layer is made of a population of five neurons for each signal channel and it is designed with a fan-in of 30 and a fan-out of 192. The hidden layer is then composed of 250 neurons, and 7 neurons are used for the output layer to represent the classes (1) dribbling in basketball, (2) playing catch with a tennis ball, (3) typing, (4) writing, (5) clapping, (6) brushing teeth and (7) folding clothes [7].

The RSNN was trained in the snnTorch framework [5] for 150 epochs employing a cross-entropy loss function on the total number of spikes for each output neuron; a fixed learning rate equal to 0.0001 and a batch size of 64 were used. The selected dataset was split into training, validation and test subsets with ratio 70:20:10. Furthermore, once defined the test set, ten random training-validation splits were produced to allow training of the model on different splits and test on unique, never seen test data. All the performed splitting accounted for stratified sampling, ensuring the same class representation as in the original dataset.

Results

Figure 1 shows the training and validation curves for both (a) loss and (b) accuracy, highlighting that even less than 150 epochs were sufficient to reach accuracy saturation and to avoid overfitting on training data. A statistical measure of the classification performance of the trained model was produced by performing 50 classification experiments on the whole test subset, which was randomly shuffled at each run to produce batches of fixed dimension but different composition. The resulting accuracy was $(94.62 \pm 0.13)\%$, and Figure 2 shows the confusion matrix produced on test data. Adopting the same strategy used in [1], an energy estimation of such model deployed on Loihi was finally performed also, which provided (1.40 ± 0.32) μJ per sample as the result of 100 single-sample inferences.

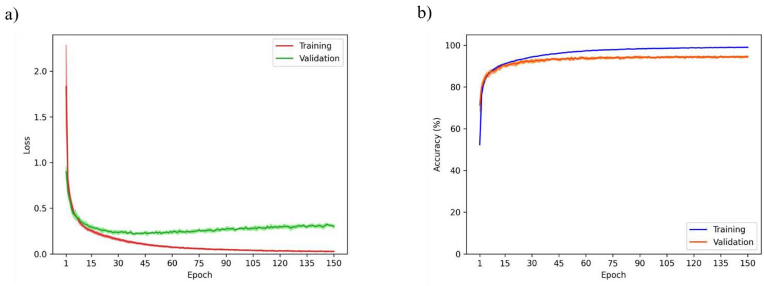


FIGURE 1

Training and validation results for **(a)** loss and **(b)** accuracy along 150 epochs. The solid lines correspond to the median values, while the shaded area is defined by the standard deviation.

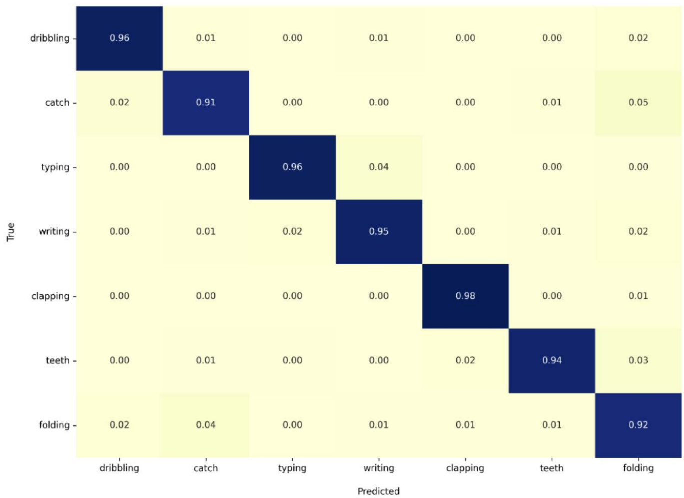


FIGURE 2

Confusion matrix obtained with the presented LIF-based, fully connected RSNN on the test set for hand-oriented, general activities from the WISDM-HAR dataset.

Conclusions

Solving tasks that involve human-related temporal data can take shape in a wide range of applications of different relevance and impact. In working environments, for instance, safety conditions can be evaluated relying on the worker behaviour through wearable devices. The proposed model of recurrent SNN employing LIF-based neurons offers a neuromorphic approach suitable for low-power, on-edge solutions, and the selected classification task with data collected from smartwatch sensors shows a possible example of point-of-care monitoring. The reported accuracy close to 95%, coupled with the estimation of a significantly small energy consumption of less than 2 μJ per sample, makes neuromorphic HAR a real candidate for online, in-situ monitoring of human activities.

References

[1] V. Fra et al., Neuromorph. Comput. Eng., vol. 2, 014006, Feb. 2022.

[2] S.F. Müller-Cleve et al., Front. Neurosci., vol. 16, 951164, Nov. 2022.

[3 M. Davies et al., IEEE Micro, vol. 38, n. 1, pp. 82-99, Jan. 2018.

[4] G. Orchard et al., 2021 IEEE Workshop on Signal Processing Systems (SiPS), pp. 254-259, Oct. 2021.

[5] J. Eshraghian. et al., arXiv:2109.12894v5, May 2023.

[6] A. Javanshir et al., Neural Computation, vol. 34, n. 6, pp. 1289–1328, Jun. 2022.

[7] G.M. Weiss et al., IEEE Access, vol. 7, pp. 133190-133202, Sep. 2019.

Mesoscale “digital twins” for motor plan cancellation in nonhuman primates

Author
Andrea Galluzzi – Natl. Center for Radiation Protection and Computational Physics, Istituto Superiore di Sanità, 00161 Roma, Italy
Nicola Alboré – PhD Program in Physics, Dept. of Physics, “Tor Vergata” University of Rome, 00133 Roma, Italy; Centro Ricerche “Enrico Fermi,” 00184, Roma, Italy
Pierpaolo Pani – Dept. of Physiology and Pharmacology, “Sapienza” University of Rome, 00185 Roma, Italy
Stefano Ferraina – Dept. of Physiology and Pharmacology, “Sapienza” University of Rome, 00185 Roma, Italy
Maurizio Mattia – Natl. Center for Radiation Protection and Computational Physics, Istituto Superiore di Sanità, 00161 Roma, Italy

Citation
Galluzzi, A., Alboré, N., Pani, P., Ferraina, S., Mattia, M. Mesoscale “digital twins” for motor plan cancellation in nonhuman primates.

Abstract
The premotor cortex plays a crucial role in adaptable behavior that is influenced by context. However, the precise mechanisms and computations that underlie this role are still widely debated. Specifically, individual premotor neurons often exhibit highly heterogeneous activation patterns making difficult to establish their specific contribution in producing the behavioral output. In this study, we investigate the activity of the premotor cortex in macaque monkeys trained to perform a motor reaching task where

the unpredictable occurrence of sensory inputs has to be taken into account in order to make a correct decision. Our findings reveal that the collective activity of the premotor cortical network can be effectively interpreted within the context of the theory of dynamical systems [1]. More specifically, the observed electrophysiological activity from *in vivo* multi-electrode recordings can be successfully replicated by an autonomous recurrent neural network working after a learning phase as an echo state network. The structural and dynamical organization of such *in silico* "digital twin" unveils the mechanisms underlying the selection and integration of relevant inputs in the task. In general, this highlights the existence of two dynamical phases tightly related to the reaction time in takin a decision: an accumulation phase driven by endogenous noise and an excitable-like dynamics likely underlying the implementation of threshold process.

Behavioral task, single-unit heterogeneity and low-dimensional latent state space
In order to identify the network dynamics of the dorsal premotor cortex (PMd) involved in generating and controlling movement, the majority of research has concentrated on tasks where actions are carefully planned and executed, as in delayed reaching tasks. However, to gain a complete understanding of the dynamics related to movement control, it is essential to investigate how movements are actively suppressed.

Indeed, inhibition plays a fundamental role in control processes, both when there are multiple possible alternatives and only one option can be chosen, and when it is necessary to sustain the current position. We derived single (SUA)- and multi-unit activity (MUA) from a multi-electrode array chronically implanted in the dorsal premotor cortex (PMd) of two macaque monkeys performing a reaching task where in a randomly selected trials a Stop signal instructs to cancel the planned movement not yet initiated Fig. 1A. As shown in Fig. 1A stop and no-stop trials (with or without the Stop signal, respectively) were intermingled on each block. Monkeys were required to touch the central target, hold the position (Hold), and wait until the Go signal (Go) instructed them to reach the peripheral target (red circles; presented either to the right or to the left; here, only one position is shown). The reaction time (RT) is the time from Go to movement initiation. In stop trials, after the Go signal, the Stop signal (central target reappearance) instructed them to stay still (correct-stop). Wrong-stop indicates trials with failed still in Fig. 1B we can see examples of single-units recorded. Spike density functions (SDFs) and raster plots for correct-stop trials (red) and latency-matched (control) no-stop trials (green) are shown for a specific SSD separately for each movement direction. Using this countermanding task, in the last years the PMd has been suggested to be a key area in motor preparation, in turn composed as a cascade of neuronal events bringing the collective activity into a preferred and stereotyped (attractor-like) state [4–5]. Such a preparatory state [2] allows neuronal population activity to vary after sensory instructions without eliciting any movement onset, thus defining an “output-null” state well separated by the inner representation of the successive movement execution [7,8]. Within this framework, movement generation is the consequence of a further change in the neuronal state, previously described as the transition from the output-null space toward the “output-potent” space [7, 8]. In this study, we employed dimensionality reduction techniques to analyze the neural activity at multiple scales (MUA and SUA). Through this analysis, we discovered the presence of a distinct subspace at the population level, referred to as the “HPA” and “PEA” (holding-and-planning and planning-and-execution axis, see Fig. 1C for SUA, and Fig. 1D for MUA). We propose that this subspace serves as a generalized form of the previously identified output-null subspace [7]. In our interpretation, the premotor neuronal activity must actively remain within this subspace to prevent the transition from the motor planning stage to execution. This dynamic control strategy expands upon the concept of the output-null subspace and complements the findings of previous studies on delayed reaching tasks.

Recurrent neural network (RNN) ‘digital twins.’

To simulate neural activity from PMd and reproduce the observed behavioral output (i.e., the reaction time distribution and the absence of movements), we designed a recurrent network (RNN) of interconnected nonlinear neurons [1] (Fig. 2A). The RNN was trained to generate a binary output on each

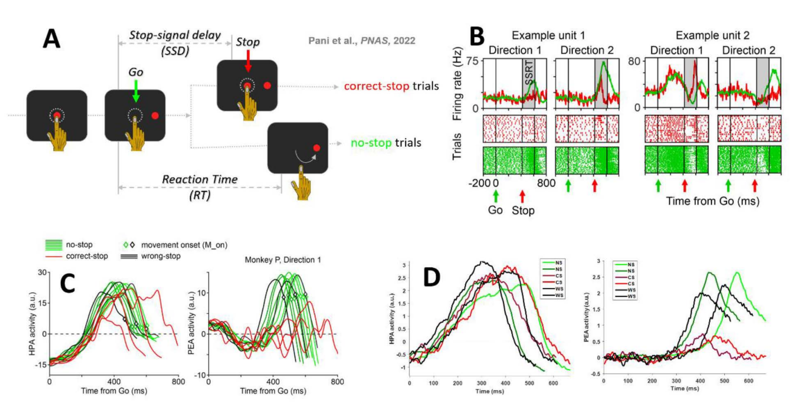


FIGURE 1

Countermanding task and neuronal correlated of motor decision in monkey PMd.

(A) Schematic representation of the countermanding task.

(B) Examples of single-units recorded. Spike density functions (SDFs) and raster plots for correct-stop trials (red) and latency-matched (control) no-stop trials (green) are shown for a specific SSD separately for each movement direction.

(C-D) "HPA" and "PEA" (holding-and-planning and planning-and-execution axis, **(C)** for SUA, and **(D)** for MUA.

trial: producing a 1 at the reaction time (i.e., when a reaching movement is initiated) or 0 when no movement is produced. At the same time, The RNN was trained to autonomously replicate the dynamics observed in the first principal components of the recorded MUAs provided as additional input (Fig. 2C Fig. 2D). Notably, we only defined what the network should achieve, with no constraints on how it should accomplish it. Thus, the solution attained by the network was not pre-programmed into its architecture. During each trial, the model network of neurons received independent sensory inputs mimicking the visual stimuli of the task (the Go, with the two possible Targets, and the Stop signals). The RNN activity was read out using a linear Perceptron, i.e., a weighted sum of the responses from all neurons in the model network. After training, the model qualitatively reproduced the

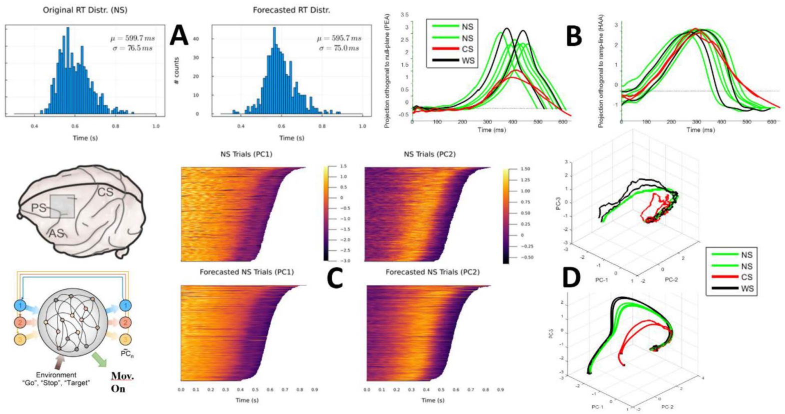


FIGURE 2

Reproducing MUA from PMd in a recurrent neural network.

(A) distribution of the experimental Reaction Times (left) and those reproduced from the reservoir (right).

(B) "HPA" and "PEA" (holding-and-planning and planning-and-execution axis, for MUAs from simulations.

(C) Activity raster of the first two principal components for no-stop trials. Trials are organized by reaction time (experimental data (above) simulated data (below)).

(D) comparison of mean trajectories in principal component space for experimental (above) and simulated (below) trials.

MUA recorded from monkey PMd (Fig. 2B, C, D), confirming that the model effectively solves the selection problem at the behavioral level (Fig. 2A).

Conclusion

In this work we have shown how recurrent neural networks (RNN) can successfully operate as "digital twins" reproducing both the behavior and the mesoscopic neural signal obtained from in vivo recording. These *in silico replica* can be used, in principle, to inspect the dynamical organization of other biological twins when a similar signal is recorded: a proxy to identify relevant features of pathological and physiological networks.

Moreover, our findings suggest that the computational processes occurring in premotor cortex arise from the coordinated dynamics of extensive neuronal populations, and can be effectively examined within the framework of dynamical systems. Importantly, the heterogeneous dynamics observed in premotor cortex responses during the integration of inputs to select a suited motor plan can be comprehensively characterized a dynamical system unfolding at the mesoscale level on two linear manifolds: the HPA and the PEA, confirming what previously found in SUA [8].

Such low-dimensional dynamics of premotor cortex stands in stark contrast to the intricate nature of single neuron responses often observed associative cortices in behaving animals. These responses exhibit a multiplexed representation of various signals pertaining to the task and decision-making processes. However, in light of our findings, these signal mixtures can be interpreted as distinct representations that can be separated at the level of the neural population.

References

[1] Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, 304, 78–80. DOI: 10.1126/science.1091277

[2] M. Weinrich, S. P. Wise, The premotor cortex of the monkey. J. Neurosci. 2, 1329–1345 (1982)

[3] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, S. I. Ryu, K. V. Shenoy, Cortical preparatory activity: Representation of movement or first cog in a dynamical machine? Neuron 68, 387–400 (2010).

[4] M. M. Churchland, K. V. Shenoy, Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. J. Neurophysiol. 97, 4235–4257 (2007).

[5] M. Mattia et al., Heterogeneous attractor cell assemblies for motor planning in premotor cortex. J. Neurosci. 33, 11155–11168 (2013).

[6] M. M. Churchland, J. P. Cunningham, A dynamical basis set for generating reaches. Cold Spring Harb. Symp. Quant. Biol. 79, 67–80 (2014).

[7] M. T. Kaufman, M. M. Churchland, S. I. Ryu, K. V. Shenoy, Cortical activity in the null space: Permitting preparation without movement. Nat. Neurosci. 17, 440–448 (2014).

[8] Pierpaolo Pani, Margherita Giamundo, Franco Giarrocco, Valentina Mione, Roberto Fontana, Emiliano Brunamonti, Maurizio Mattia, and Stefano Ferraina. Neuronal population dynamics during motor plan cancellation in nonhuman primates. PNAS 19 (28) e2122395119 (2022)

Bio-physics-based finite elements numerical simulation of an ionic emission and electrical sensing platform for neuron activity modulation: The case of calcium ions and calcium ion channels

Author
Jacopo Nicolini – DIEF, University of Modena and Reggio Emilia, Italy
Pierpaolo Palestri – DPIA, University of Udine, Italy
Luca Selmi – DIEF, University of Modena and Reggio Emilia, Italy

Citation
Nicolini, J., Palestri, P., Selmi, L. Bio-physics-based finite elements numerical simulation of an ionic emission and electrical sensing platform for neuron activity modulation: The case of calcium ions and calcium ion channels.

Introduction

Closed loop neuron activity modulation and sensing is a field of intense research [1][2][3]. Microelectronics can support these studies with the development of a new generation of ionic stimulation / electronic sensing neuromodulators for novel in-vitro experiments beyond those currently using electrical stimulation and MEAs [4][5]. The design of these new devices demands insights on the electrochemical interactions among neurons, glia and electrodes, a deep understanding of the microscopic processes leading to ionic release and electronic signal transduction, and improved modeling methods beyond the commonly adopted equivalent circuit representations [6]. Bio-physics-based finite element (FEM) numerical models offer a new and promising approach to overcome these challenges [7][8].

In this context, we investigate a prototypical electronic/iontronic hardware platform to release calcium ions and sense the neuron activity, amenable to fabrication with microelectronic technology. The system has three main components (see inset in Fig.1.b): the microelectrode sensor (connected to a readout electronics), the neuron, and an iontronic actuator made of a control electrode covered with a PEDOT-PSS conductive polymer (CP). We assume the polymer as perfectly selective to calcium ions and pre-charged either by bathing in a calcium-rich solution or by using a reservoir [9]. As a relevant case study, we focus on the interaction between calcium release and three patterns of neuron firing, related to different stages of epileptic seizures: 1) tonic firing (constant frequency of action potentials after adaptation); 2) burst firing (cluster of action potentials separated by less than 10 ms); 3) paroxysmal depolarization shift (burst followed by a sustained depolarization of the neuron membrane).

Methodology

The model was built in COMSOL Multiphysics$^{(R)}$ finite-element solver, which offers a few useful features:

- The finite element description of the system biophysics, which is a step forward in the field of neurobiology where most simulations are run at the circuit level (see e.g. NEURON$^{(R)}$), and neglect extracellular phenomena such as ion transport in the polymer and in the cleft to the neuron membrane.

- The possibility to consider both pathological and non-pathological neuron states.

For demonstration purposes, the neuron model has a single compartment formed by an ellipsoidal soma, consistently with the experiments in [10] on a culture of neurons with truncated apical dendrites. The actuation electrode covered by CP surrounds the neuron. The intra- and extra-cellular physiological electrolytes are described by the Poisson-Nernst-Planck equations. The CP is described as a two-phase system [11]. Nine ion currents represent the membrane ionic transport [10]:

$$C_m \frac{dV}{dt} = -I_L - I_{Na} - I_{NaP} - I_K - I_M - I_A - I_C - I_{Ca} - I_{AHP} - I_{extr} + I_{syn} \quad (1)$$

The transient sodium and potassium currents I_{Na} and I_K generate the action potential (AP); the persistent sodium current (I_{NaP}) and the muscarinic sensitive current (I_M) impact the amount of burstiness in the neuron. The I_A current, also involved in the firing mode, has a very limited effect in our case because the concentration of A-type channels in the soma is much lower than in the dendrites. The I_C term is involved in a faster repolarization of the membrane, while I_L is the leakage current. The calcium dynamics is mainly represented by I_{AHP}, I_{Ca} and I_{NaP}. The first is responsible for adaptation and is controlled by the calcium gradient across the membrane; I_{Ca} lets calcium ions flow into the cell when an action potential is generated. The currents are controlled by a set of gating variables, each having its dynamics and time constant [10].

The I_{NaP} gating variable (p_∞) depends on the extracellular $[Ca^{2+}]$ via a parameter (θ_p) whose relation with $[Ca^{2+}]$ is not known. Assuming a linear dependence entails that the burstiness variation due to the calcium is caused by biochemical phenomena that directly modulate I_{NaP}, while assuming a logarithmic dependence implies that it is due to the electrochemical potential across the membrane, i.e., the Nernst potential. We compared both assumptions and eventually, adopted a linear dependence from (2 mM, -41 mV) to (0 mM, -47 mV), since it qualitatively reproduces the experiments in [10]. The logarithmic dependence, instead, does not predict the AP bursts. The neuron is stimulated by means of an inward sodium current that mimics the incoming signal from a synapse, I_{syn} in (1). I_{extr}, instead, is the calcium extrusion current which represents the homeostatic mechanisms restoring the physiological calcium concentration. It is commonly assumed to be proportional to the difference between the calcium concentrations across the membrane.

Results

Figure 1.a shows that when the polymer is stimulated with a ramp voltage (black line), the calcium stored at the radical sites of the CP is released into the electrolyte. This causes a local increase of the calcium concentration that modulates the I_{NaP} current and, consistently with experiments in [10], turns the neuron to tonic firing. We also see (Fig.1.b) that the interspike interval (ISI) during calcium actuation lies in between the two extremes of tonic and burst firing. The centroid of the ISI for tonic firing neuron is almost equal to the spike period after adaptation (≅ 20 ms, orange). Moreover, the radical group concentration in the polymer (R_CP) influences the ability of the system to modulate the firing mode, demonstrating the need for polymer engineering.

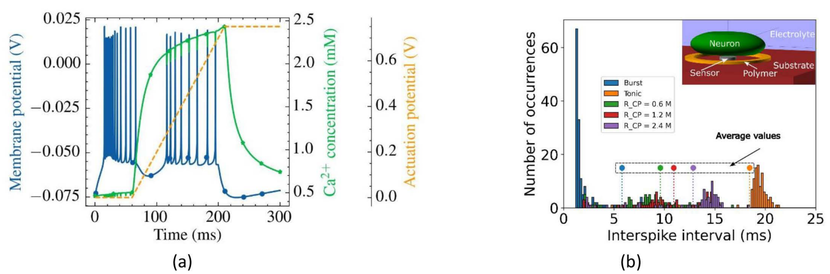


FIGURE 1

a: Example of ionic actuation: upon stimulation of the conductive polymer with a voltage ramp (orange) and the release of calcium ions next to the neuron (green) we observe a reduction in both the burstiness of the neuron and the spike frequency (blue).

b: Time distribution of the interspike intervals (ISI) extracted from long (≅ 800 ms) simulations as those in Fig.1.a without (burst mode, blue) and with (green, red, purple) calcium actuation. An increase of the doping concentration in the polymer (R_CP) results in a larger dose of calcium ions available for release, and an increased effectiveness of the actuation for the same duration of the stimulation. The ISI is compared to that of a tonically firing neuron at physiological calcium levels (tonic mode, orange). The inset shows a sketch of the actuation/sensing platform.

A lowly doped polymer is depleted of ions faster than a highly doped one; therefore, the $[Ca^{2+}]$ increase cannot be sustained for the whole duration of the actuation signal. Consequently, the concentration decays to the initial value by diffusion in the bulk, and we observe a smaller average ISI at low R_CP values.

Conclusions

A comprehensive finite element model was used to investigate an integrated ionic/electronic neuromodulation interface where fluxes of calcium ions by a pre-charged, selective polymer affect neighboring neurons. The fluxes switch the firing mode of a neuron exposed to a low $[Ca^{2+}]$ concentration, thus demonstrating a form of iontronic neuromodulation. The finite amount of ions stored in the polymer, the shape of the actuator, and the actuation waveform demand model-based optimizations as those enabled in this work to achieve the desired suppression of pathological neuronal hyperactivity. This model and the proof of concept device represent a first step towards comprehensive simulation studies supporting the development of this new ionic neuromodulation paradigm. The model is also amenable to several future extensions, e.g. aimed at including additional morphological details, (biological tissues, insulating layers, etc.), and additional biophysical complexity, (non-ideal electrolytes, convection terms, etc.).

References

[1] I. Tzouvadaki et al. 2023, doi: 10.1002/adma.202210035.

[2] M.Tambaro et al. 2022, doi: 10.1109/BioCAS54905.2022.9948694.

[3] J. Abbott, et al. 2020. doi: 10.1038/s41551-019-0455-7

[4] A. Zhang, et al. 2020 doi: 10.1016/j.nantod.2019.100821

[5] T. A. Sjöström, et al. 2018 doi: 10.1002/admt.201700360

[6] M. Zhang, et al. 2020 doi: 10.1038/s41928-020-0390-3

[7] F. Leva, et al., 2022 doi: 10.1109/SENSORS52175.2022.9967049.

[8] F. Leva et al., 2022 doi: 10.1098/rsta.2021.0013

[9] J. Isaksson et al. 2007, doi:10.1038/nmat1963

[10] Golomb et al. 2006 doi: 10.1152/jn.00205.2006

[11] Tybrandt et al. 2017 doi: 10.1126/sciadv.aao3659.

Constraint satisfaction problems solution through spiking neural networks with improved reliability: The case of Sudoku puzzles

Author
Riccardo Pignari, Vittorio Fra, Evelina Forno, Enrico Macii, Gianvito Urgese – Politecnico di Torino, Electronic Design Automation (EDA) Group, Torino, Italy

Citation
Pignari, R., Fra, V., Forno, E., Macii, E., Urgese, G. Constraint satisfaction problems solution through spiking neural networks with improved reliability: The case of Sudoku puzzles.

Introduction

Constraint satisfaction problems (CSPs) are a subset of NP-Complete problems: they belong to the NP (nondeterministic polynomial-time) class, a common example of CSP is the latin square problem in the variant of Sudoku puzzles. A possible method to find a solution is through a neuromorphic approach as shown in [1]. Specifically, a stochastic version of a Spiking Neural Network (SNN) made of Leaky Integrate-and-Fire (LIF) neurons and with a structure defined over the mathematical formulation of the CSP can be employed. Such a network describes a system that stochastically evolves towards the solution of the Sudoku puzzle, following the attractor dynamics, in the configuration space [2, 3]. Despite showing the capability of solving this and other CSP problems, [1] presents some limitations that should be addressed:

(i) each solution attempt requires a pre-set simulation time that cannot be modified once started, thus hindering the possibility of stopping the process once the solution has been found and introducing unnecessary energy consumption;

(ii) the validation method is carried out through the binning process, which consists in extracting the spikes from the neuromorphic platform and analyzing its state through an external platform which inherently implies additional energy consumption due to data preparation and transfer;

(iii) the original problem through a mapping process is encoded in the SNN network, given the complexity of some problem classes there may be limitations on the reliability of the network in modeling the initial addresses, observing a modification of these elements during the simulation.

Here we propose a fully spiking pipeline able to find a solution, validate it and stop the generation of spikes.

Methods

In our work [4], we adopted the same mapping of the Sudoku problem into an SNN as proposed in [1]. In this implementation, the variable associated with each of the 81 cells of the sudoku puzzle is represented by 9 populations of neurons, one for each possible value, in a winner-take-all configuration. The constraints are modelled through lateral-inhibition by connecting different populations across cells. The resulting structure defines a sparsely connected graph which encloses the formulation of the original problem in the number of neurons and their connections. In order to address the above mentioned issues of [1] we implemented two major changes. Hereinafter, we refer to the original SNN as in [1], being the one directly in charge of exploring the configuration space, as the *CSP-Solver*, and to our changes as the *Clue-Hold* and the *Spike-Val* subnetwork.

The *Spike-Val* subnetwork is introduced to deal with issues (i) and (ii), and it is composed of 3 blocks with different tasks:

- *Block-1*: it verifies that the constraints on the cell's values are met, and it consists of two levels:

1. It has a structure like that of the *CSP-Solver* with a reduced number of neurons per population, which is used to clear the SNN output of the background noise.

2. It allows us to verify that the 3 types of constraint, namely on the row, on the column and within the subgrid, are satisfied.

- *Block-2*: with a structure made of a single layer, it verifies the solution, interrupts the activity of the main network and activates *Block-3).*

- *Block-3*: it holds the activity state of the solution found by the main network.

The *Clue-Hold* tackles instead issue (iii) by implementing unidirectional lateral-inhibition from the cells corresponding to the clues to the empty ones. Such symmetry break ensures that the initial clues cannot be affected by the activity of the populations they must hinder from assuming specific values along a row, a column or within a subgrid.

Conclusion

To verify our implementation, we focused on the class *easy* of Sudoku puzzles, using the one reported in [1] and two additional ones of the same class [5] for 300 solution trials for each of them. All of the 900 trials have been performed with both the original implementation [1] and our proposed pipeline. In *Figure 1a,* the two approaches are compared in terms of solutions found, and validated, expressed as a percentage of the total set of experiments. The gain of more than 9% we achieved can be ascribed to the *Clue-Hold*, which ensures that the original formulation of the CSP is held during the overall simulation. A direct and quantitative evaluation of the impact of *Clue-Hold* is shown in *Figure 1b* where more than 11% of the experiments carried out with the original implementation were affected

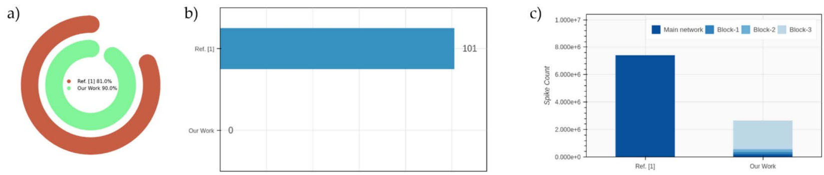


FIGURE 1
Results from 900 trials. **(a)** Percentage of solutions found and successfully validated. **(b)** Number of clues change occurring during the network activity. **(c)** Median value of the total spike count.

by such an issue, which was instead completely fixed with our SNN solver. *Figure 1c* reports the median of the spiking activity produced, where all the contributions from the different portions of the network are highlighted. Despite having additional neurons involved in the dynamic evolution of the system, our implementation reduces to about 3× the total number of spikes, such a result arises from the possibility, given by the *Spike-Val* subnetwork, of deactivating the main network once a solution is found. Starting from [1], we therefore have shown that the solution to a specific class of CSPs as that of Sudoku puzzles can be found and validated in an efficient, reliable and fully spiking way.

References

[1] G. A. Fonseca Guerra et al., Front. Neurosci., vol. 11, p. 714, Dec. 2017.

[2] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A., vol. 79, n. 8, p. 2554-2558, Apr. 1982.

[3] G. Pedretti et al., IEEE Int. Symp. Circuits Syst., p. 1-5, Oct. 2020.

[4] R. Pignari et al., *under submission*.

[5] T. Mantere et al., IEEE CEC, p. 1382-1389, Sep. 2007.

Resilience of an inferential neuromorphic circuit

Author
Mirco Tincani – Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy
Giulia Maria Boiani – Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy
Lorenzo Benatti – Department of Engineering 'Enzo Ferrari', University of Modena and Reggio Emilia, Via P. Vivarelli 10, 41125 Modena, Italy
Tommaso Zanotti – Department of Engineering 'Enzo Ferrari', University of Modena and Reggio Emilia, Via P. Vivarelli 10, 41125 Modena, Italy
Francesco Maria Puglisi – Department of Engineering 'Enzo Ferrari', University of Modena and Reggio Emilia, Via P. Vivarelli 10, 41125 Modena, Italy ; Centre for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy
Giuseppe Pagnoni – Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy
Jonathan Mapelli – Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy ; Centre for Neuroscience and Neurotechnology, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy
Daniela Gandolfi – Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy

Citation
Tincani, M., Boiani, G.M., Benatti, L., Zanotti, T., Puglisi, F.M., Pagnoni, G., Mapelli, J., Gandolfi, D. Resilience of an inferential neuromorphic circuit.

Introduction

According to the Bayesian Brain theory, the central nervous system can be conceived as a prediction machine that continuously generates expectations about upcoming sensory data based upon an internal model of the world. This proactive modality affords unique performance capabilities to biological agents compared to mere feedback control loops. In such perspective, the Active Inference (AI) framework predicates that the processes of Bayesian updating unfold under the imperative of minimizing a tractable proxy of surprise – i.e., the discrepancy between expected and actual observations – named variational free energy [1]. It has been recently shown that the AI framework can be deployed to model neural networks where each neuron functions as a Bayesian inference agent [2,4]. Such networks, endowed with a biologically plausible architecture, showed the emergence of autonomous learning as evidenced by persistent changes of synaptic efficacies. Here, we present a model of an inferential network inspired by the dopamine reward circuit, one of the most phylogenetically conserved brain pathways responsible for our survival. The proposed architecture is a simplified version of the biological circuit [3], with two parallel pathways converging onto the same readout neurons mediating the processing of conditioned (CS) and unconditioned stimuli (US), respectively, see Fig1A.

Results

Excitatory and inhibitory synapses automatically adjusted their strength following solely the free energy minimization (Fig1B). To quantify the effect of conditioning, the total number of spikes (spike area) was computed in the temporal window corresponding to conditioned stimulus. The spike histogram of the activity of readout layer, corresponding to the Nucleus Accumbens (NAc), was divided in two equal time-windows, one in the first half and the other in the second half of the window. This procedure allowed the identification of the early response of the NAc upon the combined presentation of CS and US stimuli (USCS) compared to the CS or US stimuli alone (see Fig2A). Moreover, the expression of the conditioning mechanism was strengthened by the introduction of neuromodulatory elements, such as the dopaminergic input from the dopaminergic neurons to the medium

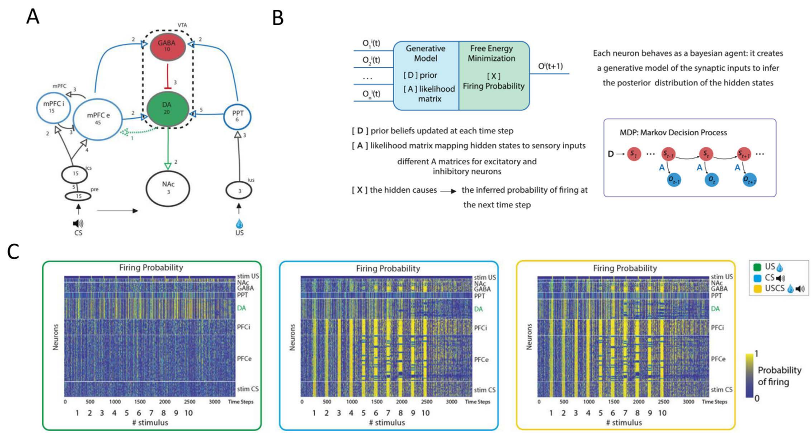


FIGURE 1

(A) The dopamine reward circuit with two parallel pathways converging onto the same readout neurons mediating the processing of conditioned and unconditioned stimuli (VTA: ventral tegmental area, GA: GABAergic neurons, DA: dopaminergic neurons; mPFC: medial prefrontal cortex, e: excitatory, i: inhibitory; PPT: pedunculopontine nucleus; NAc: nucleus accumbens; ics: stimulus from CS (Conditioned Stimulus) pathway; ius: stimulus from US (Unconditioned Stimulus) pathway). Arrows represents excitatory connections and connectors with flat end represents inhibitory connections. Numbers inside the areas are the numbers of neurons assigned to each area, based on the anatomical literature. Numbers on the arrows represent the connectivity divergence ratio. **(B)** Schematic representation of a single neuron acting as a Bayesian agent. Each neuron encodes with its firing probability (X), an inference about the hidden states of neurons generating its inputs through the iterative minimization of the free energy. The firing probability X encodes the update of the prior beliefs about the presynaptic hidden states **(D)**. At time t, the ith neuron receives inputs from different presynaptic neurons (o_i) based on likelihood matrix **(A)**. Every neuron can be considered presynaptic for other neurons at a subsequent time step. **(C)** The panels show the firing probability (X) for each neuron as a function of timesteps (panel 1: US; panel 2: CS; panel 3: USCS). On the x axis the timesteps and the stimuli numbers are shown. On the y-axis each row shows the X for each neuron. White lines separate neurons belonging to different pools, labeled on the right of each panel.

prefrontal cortex, which positively affected the associative learning outcome. To test network resiliency to perturbation, the performance of the network was evaluated by introducing a randomly distributed noise that altered the computation of the equivalent of the firing rate, as shown in Fig2B.

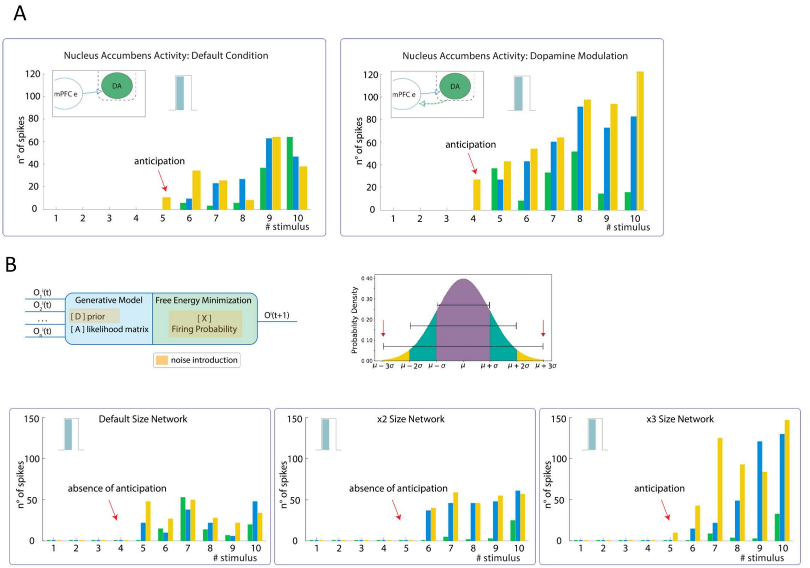


FIGURE 2

(A) Histograms represents the number of spikes generated by all the NAc neurons within the first half of the stimulus time window in absence (left panel) and presence (right panel) of dopamine modulation, under the different stimulation protocols (US: green, CS: blue, USCS: yellow). The anticipation of the NAc response in the combined stimulation protocol is highlighted. This anticipation is even more evident in presence of dopamine modulation (right panel). **(B)** To simulate the noise impact on X and D update a randomly picked value was taken from a normal distribution and added to the variable. The normal distribution was chosen to have a standard deviation 1/3 of 10% of the variable value, corresponding to a maximum error of +/-10% on X and D. Histograms show NAc activity in presence of noise upon network rescaling from default size to two times or three times the default size. The recovery of the task was obtained starting from a network three times the default network size.

We observed that, upon the introduction of noise, the associative learning was abolished at the default size network, while the task recovered when the size of the network was progressively increased.

Conclusions

The proposed approach allows to model artificial neural networks relying on neurons behaving as if they were inferring the causes of their inputs. The framework can be generalized to the modelling of networks inspired by the hierarchical organization of real brain circuits. Interestingly, the introduction and propagation of noise within the circuit can be compensated by network rescaling suggesting the role of neuronal abundancy observed in biological systems as a strategy to achieve noise tolerance. We also envisage the possibility to replicate in hardware the proposed network of inferential neurons by designing a cost-effective reinforcement learning chip using recently developed ultra-low power hardware neurons with high-performance memristive synapses. Embedding this revolutionary technology in commercial products holds a great potential for decreasing the need for massive hardware and runtime in data centres for computationally intensive tasks, realizing significant advantages in terms of economic and carbon emissions costs.

References

[1] Friston et al., *J. Physiol.* 100, 70–87, 2006

[2] Gandolfi D. et al., J Neural Eng.19(3), May 2022

[3] Deperrois N. et al., Front Neural Circuits. 8;12-116, Jan. 2019

[4] Palacios ER. et al., Sci Rep. 9(1):6412, Apr. 2019

RRAM-based Binarized neural networks inference accelerator exploiting multiple in-memory computing paradigms

Author
Tommaso Zanotti, Paolo Pavan, Francesco Maria Puglisi – Dipartimento di Ingegneria "Enzo Ferrari," Università degli studi di Modena e Reggio Emilia, Via P. Vivarelli 10/1, 41125 Modena, Italy

Citation
Zanotti, T., Pavan, P., Puglisi, F.M. RRAM-based Binarized neural networks inference accelerator exploiting multiple in-memory computing paradigms.

Introduction

Brain-inspired computing paradigms have sparked remarkable technological advancements in neuroscience, electronics, and computer science. This progress stems from the growing need for intelligent agents deployed at the edge of computing infrastructure, where energy efficiency and timely decision-making is crucial. Although artificial neural networks only capture a subset of the brain's characteristics, they have emerged as effective solutions for classification tasks. Binarized neural networks (BNN) [1], which encode weights and activations with a single bit, offer resource-efficient implementations for low-power edge devices. However, the power limitations of these devices necessitate dedicated ultra-low-power hardware accelerators for BNNs.

In-memory computing architectures and emerging nonvolatile memory nanodevices inspired by the brain's co-localization of memory and processing are key enabling technologies. Resistive random access memory (RRAM) technologies and accelerators based on Logic-in-memory (LiM) and mixed-signal vector-matrix multiplication (VMM) frameworks show promise for ultra-low-power BNN inference accelerators. While these accelerators are typically optimized for specific functions, resource-constrained devices would benefit from enhanced hardware reconfigurability. Recently, a hardware accelerator enabling the implementation of multiple in-memory computing paradigms on the same memory array has been proposed [2], [3].

In this work, the performance evaluation of the BNN inference hardware accelerator enabling the coexistence of LiM and analog VMM frameworks on the same crossbar array is conducted using the UNIMORE RRAM physics-based compact model [4], calibrated on a TiN/HfO_x/AlO_x/Pt RRAM technology from the literature [5]. The estimated performance is benchmarked on a classification task on the MNIST dataset implemented on a fully connected neural network. This assessment provides insights into the capabilities and effectiveness of the proposed hybrid architecture.

The hybrid RRAM-based BNN accelerator

The hybrid RRAM-based BNN accelerators combine the smart IMPLY (SIMPLY) LiM framework introduced in [6] with the mixed-signal VMM accelerator described in [7]. The SIMPLY LiM framework enables the efficient acceleration of logic operations computed on the results of the VMM operations (i.e., intermediate result accumulation, and activation function implementation). The mixed-signal VMM accelerator enables the parallel computation of the sum of products between input activations and neuron weights.

The SIMPLY LiM framework [6], offers an efficient solution for implementing the material implication logic on RRAM devices, and its core functional circuit is shown in Fig. 1c. Its operation is based on two core logic function the IMPLY, and the FALSE that together form a complete logic group.

The IMPLY operation is split into two steps, i.e., a read step and a conditional programming step, see Fig. 1d. During the read step, small voltage pulses (~100 mV) are applied to the input RRAM devices, allowing discrimination between different input configurations using a comparator with a threshold voltage (V_{TH}). In the conditional programming step, the programming voltage (i.e., V_{SET}) is selectively delivered to the output device based on the comparator's output. Circuit simulations with the compact model confirm the efficacy of this approach in mitigating variability and random telegraph

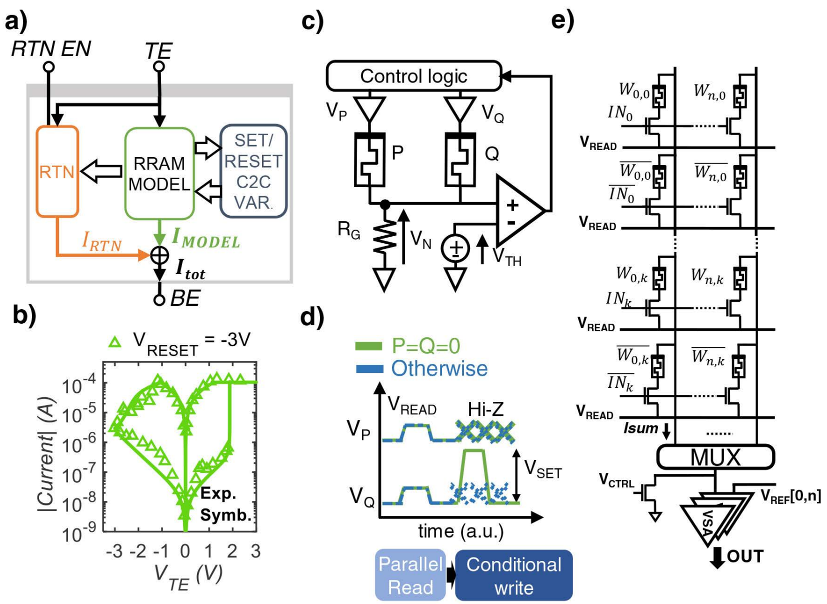
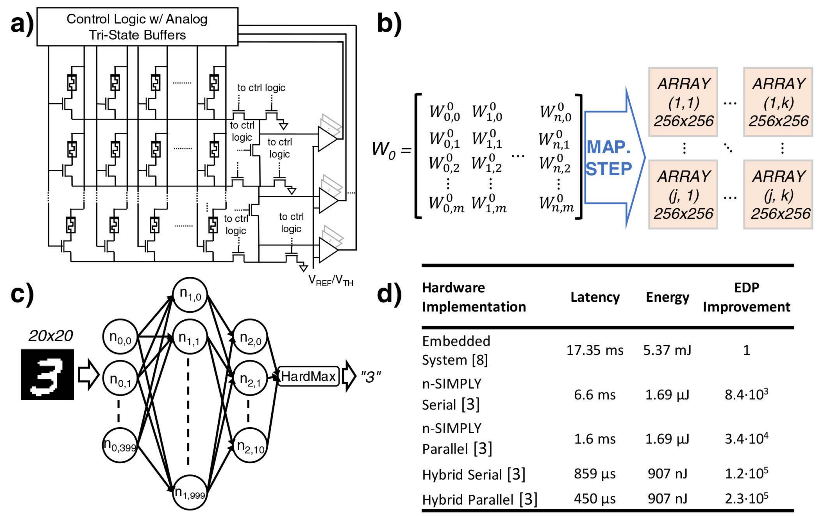


FIGURE 1

a) Illustration of the functional block diagram of the compact model, showcasing the internal representations of the device nonidealities. **b)** Experimental and simulated quasi-static IV of a TiN/HfO_x/AlO_x/Pt RRAM device. **c)** Core circuit of the SIMPLY framework. **d)** The signals employed for executing an IMPLY operation within the SIMPLY framework. V_{SET} is delivered to Q when the input configuration P=Q=0 is detected, while the voltage drivers remain in a high impedance (Hi-Z) state for all other input combinations. **e)** RRAM-based BNN vector-vector multiplication accelerator. Two RRAM devices in adjacent columns encode a single weight of an artificial neuron.



Hardware Implementation	Latency	Energy	EDP Improvement
Embedded System [8]	17.35 ms	5.37 mJ	1
n-SIMPLY Serial [3]	6.6 ms	1.69 μJ	$8.4 \cdot 10^3$
n-SIMPLY Parallel [3]	1.6 ms	1.69 μJ	$3.4 \cdot 10^4$
Hybrid Serial [3]	859 μs	907 nJ	$1.2 \cdot 10^5$
Hybrid Parallel [3]	450 μs	907 nJ	$2.3 \cdot 10^5$

FIGURE 2

a) Circuit diagram depicting the hybrid in-memory computing accelerator that enables the simultaneous operation of the SIMPLY and analog BNN VMM frameworks on a single RRAM array. The shaded VSAs indicate potential utilization of flash ADCs within the architecture. **b)** Mapping operation of the neural network parameters onto RRAM arrays. The number of arrays (represented as j · k) varies depending on the number of inputs (m) and outputs (n) within each layer of the neural network. **c)** Implemented fully connected neural network solving a classification task on the MNIST dataset. **d)** Benchmark of the performance of different SIMPLY based accelerators against a state-of-the-art microcontroller-based implementation. For the IMPLY-based accelerators a 500 MHz clock frequency was considered. "Serial" and "Parallel" refer to accelerator implementations which exploit no computing parallelism and the maximum computing parallelism, respectively.

noise (RTN) effects. The execution of complex logic functions requires sequences of IMPLY and FALSE operations along with an adequate number of devices.

The core functional circuit of the BNN VMM accelerator inspired by [7], is shown in Fig. 2e. A single weight of a neuron is encoded into the resistive state of two RRAM devices programmed in complementary resistive states located in adjacent rows. Weights belonging to the same neuron are located on the same columns of the array. Thus, by delivering the input activations to the rows of the array, the vector matrix multiplication is computed in a single step, and the result of the operation is encoded in the current flowing in each column, which is then converted into a voltage using a FET device operating in linear region, and digitized by means of a flash ADC implemented with multiple VSAs.

The hybrid architecture merging the two approaches is shown in Fig. 2a. IMPLY and FALSE operations can be performed on devices located on the same row and different columns in the array by appropriately driving the select transistors and transistors in the array periphery. Conversely, to execute BNN VMM operations, weights associated with the same neuron are encoded into two adjacent rows, input activations are delivered to the array columns, and the currents encoding the results of multiplications flow in the rows of the array.

Performance benchmark
The hybrid architecture's performance was evaluated on an MNIST dataset for a BNN inference task and compared to a conventional embedded system implementation [8]. The neural network used has a single hidden layer with 1000 neurons and 10 output neurons, trained using the DoReFa-Net algorithm [9]. The trained network achieves an accuracy of 91.4% [3]. The performance assessment considered computations executed on a suitable number of 256x256 1T1R memory arrays, as depicted in Fig. 2b. Additionally, a reference implementation performing the VMM operation in the SIMPLY framework was simulated. Simulation results (Fig. 2d) demonstrate that the hybrid accelerator surpasses both the embedded system implementation (EDP improvement > 10^5) and the fully SIMPLY implementation (EDP gain > 3.5).

Conclusions
This work focused on the implementation of a hybrid architecture that integrates LiM and BNN VMM frameworks on a single hardware platform.

The EDP evaluations of this RRAM-based BNN inference accelerator were compared to a state-of-the-art embedded system implementation, revealing significant energy enhancements. This hardware accelerator offers both high reconfigurability and energy efficiency, making it a valuable solution for future ultra-low power hardware.

References

[1] Hubara et al., NIPS, vol. 29, p. 4107, 2016.

[2] T. Zanotti et al., JLPEA, vol. 11, n. 3, Sep. 2021.

[3] T. Zanotti et al., IntechOpen, Neuromorphic Computing, Mar. 2023.

[4] T. Zanotti et al., Microelectron. Eng., vol. 266, p. 111886, Sep. 2022.

[5] S. Yu et al., VLSI-TSA 2011, Hsinchu, p. 1-2, Apr. 2011.

[6] T. Zanotti et al., IEEE JEDS, vol. 8, p. 757-764, Apr. 2020.

[7] S. Yin et al., IEEE TED, vol. 67, n. 10, p. 4185–4192, Oct. 2020.

[8] B. McDanel et al, EWSN 2017, Uppsala, p. 168–173, Feb. 2017.

[9] S. Zhou et al., arXiv, Feb. 2018

Where scientists empower society

Creating solutions for healthy lives on a healthy planet

frontiersin.org

Why publish with us?

Open access

All Frontiers journals are fully open access, meaning every research article we publish is immediately and permanently free to read.

Peer review

Our collaborative peer review is handled by experts in the field who objectively certify the quality, validity, and rigor of research.

Technology

With the latest custom-built technology and artificial intelligence, we're revolutionizing the way research is published, evaluated, and communicated.

Impact

We are the third most-cited publisher with 1.4 billion article views and downloads – reflecting the power of research that is open for all.

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Swizerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact