Ultra-Compact, Entirely Graphene-based Nonlinear Leaky Integrate-and-Fire Spiking Neuron

Wang, H.; Cucu Laurenciu, N.; Jiang, Y.; Cotofana, S.D.

DOI
10.1109/ISCAS45731.2020.9181092

Publication date
2020

Document Version
Accepted author manuscript

Published in
ISCAS 2020: IEEE International Symposium On Circuits & Systems

Ultra-Compact, Entirely Graphene-based Nonlinear Leaky Integrate-and-Fire Spiking Neuron

H. Wang, N. Cucu Laurenciu, Y. Jiang, S.D. Cotofana
Computer Engineering Laboratory, Delft University of Technology, The Netherlands.
{H.Wang-13, N.CucuLaurenciu, Yande.Jiang, S.D.Cotofana}@tudelft.nl

***Abstract*—Designing and implementing artificial neuromorphic systems, which can provide biocompatible interfacing, or the human brain akin ability to efficiently process information, is paramount to the understanding of the human brain complex functionality. Energy-efficient, low-area, and biocompatible artificial neurons are key ubiquitous components of any large scale neural systems. Previous CMOS-based neurons implementations suffer from scalability drawbacks and cannot naturally mimic the analog behavior. Memristor and phase-changed neurons have variability-induced instability drawbacks, and usually rely on additional CMOS circuitry. However, graphene, despite its ballistic transport, inherently analog nature, and biocompatibility, which provide natural support for biologically plausible neuron implementations has only been considered for Boolean logic implementations. In this paper, we propose an ultra-compact, all graphene-based nonlinear Leaky Integrate-and-Fire spiking neuron. By means of SPICE simulations, we validate its basic functionality and investigate the output spikes response under stochastic noisy input spike trains with a variable firing rate, from 20 to 200 spikes per second. Simulation results indicate neuron robustness to noisy scenarios, and neuronal output firing regularity. The small area and the low energy consumption, due to $200\,\mathrm{mV}$ supply voltage operation, can benefit the implementation of large scale neural networks, and the biologically plausible operating conditions (e.g., $2\,\mathrm{ms}$ and $100\,\mathrm{mV}$ spike duration and amplitude), can promote the interfacebility of graphene-based artificial neurons with biological counterparts.**

***Index Terms*—Neuromorphic Computing, Integrate-And-Fire Neuron, Graphene, GNR.**

I. Introduction

The human brain's unique, outstanding properties (e.g., energy efficiency, suitability for complex task solving, real-time reaction and highly parallel information processing ability) make it a powerful high performance computing system, which promotes the development of novel biologically-inspired computation paradigms (e.g., neuromorphic computing) targeted to understand the brain intrinsic operational principles and obtain biological brain-alike computation abilities.

Since the nervous system, which supports the human brain complex functionality, comprises billions of neurons, it makes the design and implementation of large-scale neuromorphic computing systems an extremely challenging task. State-of-the-art CMOS-based artificial neurons use complex CMOS circuitry and have a relatively high power consumption [1], [2], which limit the complexity, scalability, and energy efficiency of achievable neuromorphic system implementations. Besides, CMOS-based neurons cannot intrinsically mimic the analog behavior of biological neurons. Recently, emerging resistive switching memory devices [3] attracted interest and have been utilized in spiking neurons implementations [4], [5], due to their analog behavior, ability to restore the state memory, and good scalability. However, they suffer from resistive state temporal and spatial variability and undesired stochastic behavior, which may cause neuromorphic systems instability. Artificial neurons based on the phase-change devices were also proposed as an alternative for scalable neuromorphic systems [6], [7] as their accumulation property can provide a proper electronic mimicry of spiking neurons membrane potential dynamics. However, phase change neuron implementations require additional CMOS circuitry to emulate the neuron functionality and rely on externally generated auxiliary signals that control the basic functionality of phase-change devices. They also operate at relatively high voltages, which impede the implementation of energy efficient neuromorphic systems.

Graphene, [8] has lately emerged as one of the most promising materials for nanoelectronics, as it exhibits ballistic transport, ultimate thinness, an inherently analog nature, and is flexible and biocompatible. Due to its properties graphene transistor-based logic, which follows the traditional CMOS design style has been proposed in [9] [10], while alternative approaches towards gate realizations departing from the switch-based mainstream have been introduced in, e.g., [11], [12]. Moreover, as graphene is biocompatible and can model complex functionality within a single Graphene Nanoribbon (GNR), GNR-based synapses have been proposed in [13].

This paper investigates graphene's potential towards low cost and energy effective implementations of biologically plausible neurons. Specifically, we propose an all graphene-based ultra-compact and low voltage neuron, which is able to emulate the essential features of spiking neurons, including the membrane potential accumulation, the firing event, the refractory effect, and the output spike generation. The proposed neuron is operated with voltage ranges akin to those of biological neurons, which makes it a good candidate for biologically plausible utilization scenarios. The neuron consists of 6 GNR-based devices controlled via top-gate voltages, one of them emulating the membrane potential dynamics, and the remaining 5 generating the necessary control signals as well as the output spikes. We validate the basic nonlinear Leaky Integrate-and-Fire (LIF) neuron functionality with periodic input spike trains. We further evaluate the neuron output spike response when subjected to noisy stochastic input. All experiments are carried out by means of SPICE simulation. The obtained results indicate robustness to neuronal signals variability, and regular output firing rate statistics with a

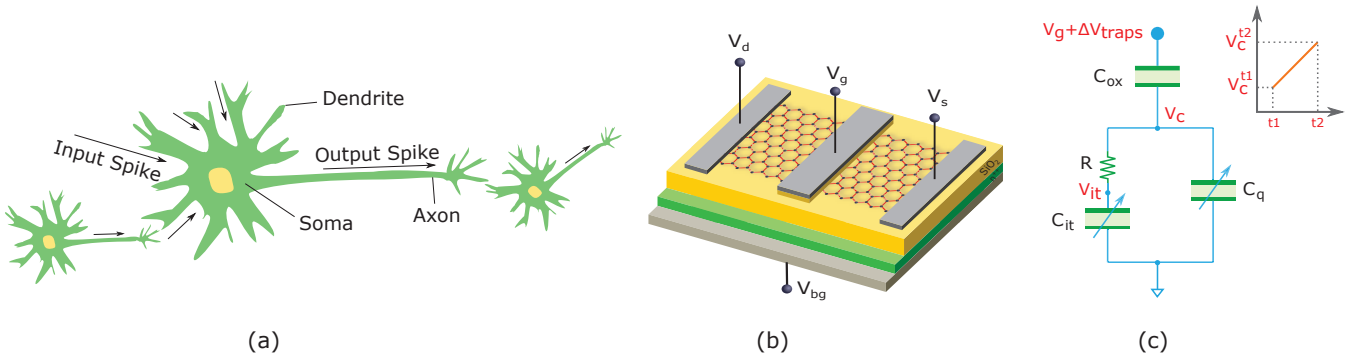

Fig. 1: Graphene-based device for artificial neurons: (a) Neuron structure, (b) Basic GNR-based device, and (c) Equivalent traps-aware capacitive circuit.

slowly decreasing trend and < 1 interspike interval variation coefficient, when increasing the input firing rates from 20 to 200 spikes per second. For all simulation, we used spike duration and amplitude of $2\,\mathrm{ms}$ and $100\,\mathrm{mV}$, respectively, which are comparable to those observed in biological neurons. Note that, the low area footprint (GNR-based device area of max. $36\,\mathrm{nm}^2$) and low energy consumption ($200\,\mathrm{mV}$ supply voltage) prove the suitability of our proposal for large-scale integration.

The remaining of this paper is organized as follows: Section II explains the basic concepts of nonlinear leaky integrate-and-fire neuron, and introduces the basic building block for graphene-based neurons. In Section III we describe the design of the proposed graphene-based neuron and explain its operation principle. Section IV presents simulation results and Section V concludes the paper.

II. Background

In this section we introduce the neuron structure, the nonlinear Leaky Integrate-and-Fire (LIF) model, and the fundamental building block for the proposed graphene-based neuron.

As illustrated in Figure 1 (a), a neuron comprises: (i) a soma, which is the neuron's cell body where the main neuronal dynamics occur (e.g., membrane potential evolution, spike generation, and refractory effect), (ii) dendrites, which connect the neuron with other neurons, receive and process input spikes, and generate neuronal input trains to the soma, and (iii) an axon, which is a long nerve cell, that transmits the output spike generated by the soma to neighbouring neurons. Various neuron models are proposed to describe the behavior of biological neurons, among which the Integrate-and-Fire neuron model attracts particular interest, due to its low complexity that makes it easy to analyze neuronal behavior while being able to capture the essential properties of biological neurons. In a standard nonlinear Leaky Integrate-and-Fire (LIF) neuron, the membrane potential evolution is in line with the following equation:

$$du/dt = F(u) + G(u) \cdot I, \tag{1}$$

where u is the membrane potential, $F(u)$ denotes a voltage-dependent leak term, and $G(u)$ is the voltage-dependent input resistance, which accounts for the membrane potential accumulation due to the neuron input current I. The neuronal dynamics of a nonlinear LIF neuron can be described via: (i) an integration process, when the membrane potential u increases continuously (starting from the resting potential u_{rest}) due to input spikes contributions, (ii) a firing event that generates a neuron output spike when the membrane potential reaches a certain firing threshold θ and then immediately resets to a new value $u_{\mathrm{reset}} < u_{\mathrm{rest}}$, and (iii) a refractory period, during which the neuron cannot fire, and the membrane potential is reset to the resting potential u_{rest}.

To implement the nonlinear LIF neuron with graphene-based devices, we rely on the basic building block, illustrated in Figure 1 (b), which consists of a monolayer Graphene Nanoribbon (GNR) located above an insulating material and a doped substrate that serves as back-gate. The GNR works as a conduction channel when applying a bias voltage V_{d}-V_{s} between the source and drain terminals. The GNR conductance can be modulated by changing the graphene sheet geometry and the contacts topology as well as by means of external voltages via the top/back gates. Figure 1 (c) illustrates the equivalent capacitive circuit of the device in Figure 1 (b), where C_{ox} is the top gate oxide capacitance, C_{q} the GNR quantum capacitance, and C_{it} the capacitance caused by interface traps. Note that, it was experimentally observed that GNR devices inherently exhibit near-interface traps [8], which will trap/release charges via capacitance C_{it} in an analogue manner with the membrane potential accumulation. When applying a top gate voltage, V_{g}, charge transfer to/from graphene to the interface traps causes an equivalent shift of V_{g}, with a quantity denoted as $\Delta V_{\mathrm{traps}}$ [14]. Considering a piece-wise linear V_{g}, when the GNR surface potential V_{c} changes from V_{c}^{t1} at time moment $t1$ to V_{c}^{t2} at time moment $t2$, the interface traps charges can be obtained as:

$$Q_{\mathrm{it}}(t) = C_{\mathrm{it}} \cdot [(V_{\mathrm{c}}^{t1} + \alpha \cdot t - \alpha \cdot \tau) + e^{-\frac{t}{\tau}} \cdot (\alpha \cdot \tau - V_{\mathrm{c}}^{t1} + V_{\mathrm{it}}^{t1})], \tag{2}$$

where V_{it}^{t1} is the accumulated voltage drop on C_{it} at time moment $t1$, τ is the trapping/detrapping time constant, and α is the V_{c} ramp slope from $t1$ to $t2$. Thus with a single graphene device, the membrane integration features are naturally captured by the interface charge trapping/detrapping phenomena.

III. Graphene-based Neuron

In this section we introduce the proposed graphene-based nonlinear LIF neuron circuit and describe its operation.

As illustrated in Figure 2 (a), the graphene-based neuron comprises six GNR-based devices, which can be divided into 2 blocks: the integrate-and-fire block, which mimics the membrane potential dynamics and the output block, which generates the output spikes. To aid the explanation, we make use of the basic operation example depicted in Figure 2 (c). The neuron kernel is GNR^2_{up}, which captures the membrane potential dynamics via its conductance. Due to the GNR inherent interfacial traps, electrical charges proportional to the GNR applied voltages can be accumulated or released. Starting from the membrane resting level, such behavior can be observed until reaching the membrane firing threshold, at which point, there is a maximum accumulation of charges (which corresponds to a maximum conductance value). We denote this integrate-and-fire region as Stage I. Further, to emulate the membrane potential reset, most of the trapped charges need to be released, situation which happens only when the GNR^2_{up} top gate voltage is very small (e.g., $\approx 100\times$ smaller V_{in}) - Stage II. Then, during the refractory period, a gradual accumulation of charges should follow in order to reach the membrane resting level, situation which is achieved when applying a slightly bigger top gate voltage (but smaller than the membrane resting level), e.g., $\approx 2\times$ smaller V_{in} - Stage III. The sub-circuit composed out of GNR^1_{up} and GNR^1_{dn}, receives the neuronal input spike train V_{in} and controls the top gate voltage of GNR^2_{up} via V_{internal} (it either directly outputs the neuron input V_{in} during Stage I or a magnitude down-scaled neuron input, i.e., $\approx V_{\text{in}}/100$ during stage II and $\approx V_{\text{in}}/2$ during Stage III). The output block containing GNR^3_{up} and GNR^3_{dn} devices generates the neuron output spike V_{out}.

As illustrated in Figure 2 (c), initially, V_{internal} follows V_{in} and V_{stage} values are afferent to Stage I. When $V_{\text{internal}}+\Delta V_{\text{traps}}$ reaches the firing threshold, V_{stage} switches to Stage II and V_{internal} becomes equal to $V_{\text{in}}/100$. Charges are depleted, the membrane potential resets, and an output spike V_{out} is triggered. When $V_{\text{internal}} + \Delta V_{\text{traps}}$ reaches the voltage value which corresponds to the end point of the neuron input spike V_{in}, V_{stage} transitions to Stage III, and V_{internal} is generated equal to $V_{\text{in}}/2$. When $V_{\text{internal}} + \Delta V_{\text{traps}}$ reaches a fixed out of refractory threshold voltage level, V_{stage} switches back to Stage I, and the neuron activity resumes.

To obtain the desired GNR topologies, we performed a design space exploration, by changing the GNR geometry, and the width and position of the top-gate, such that for every up/down pair of GNRs the in-between voltage follows the aforementioned behavior. The in-between voltage can be calculated by using a voltage divider $V_{\text{DD}} \cdot G_{\text{up}}/(G_{\text{dn}} + G_{\text{up}})$, where G_{up} and G_{dn} represent the conductance of GNR_{up} and GNR_{dn}, respectively, and $V_{\text{DD}} = 0.2\,\text{V}$ denotes the supply voltage. Figure 2 (b) depicts the obtained GNR topologies, with $W \times L$ dimensions $23a \times 30\sqrt{3}a$ and $35a \times 30\sqrt{3}a$ for GNR^1_{up} and GNR^1_{dn}, respectively, $23a \times 30\sqrt{3}a$ and $29a \times 25\sqrt{3}a$ for GNR^2_{up} and GNR^2_{dn}, and $35a \times 30\sqrt{3}a$ for both GNR^3_{up} and GNR^3_{dn}, where $a = 0.142\,\text{nm}$ is the distance between 2 adjacent carbon atoms.

IV. Simulation Results

In order to model the graphene electronic transport properties we make use of the atomistic-level tight binding Hamiltonian matrix to describe the interactions between carbon atoms and external graphene potentials, the Non-Equilibrium Green Function (NEGF) formalism to solve the Schrödinger equation, and the Landauer-Büttiker formula to derive the GNR current and conductance [15]. As interface traps profile, we employed a trapping/detrapping time constant of $1.6\,\text{ms}$ and an interface trap density of $2.363 \cdot 10^{13}\ \text{cm}^{-2}(\text{eV})^{-1}$ [16] , [17].

The neuron circuit was functionally validated and evaluated by means of SPICE simulation in Synopsys HSPICE [18]. In order to preserve the GNRs physical simulation accuracy degree, we developed a Verilog-A SPICE compatible generic model, which relies on look-up tables containing GNRs conductance values for varying input profiles, which are obtained with aforementioned atomistic-level formalization. For instance, to calculate the GNR^2_{up} conductance for a certain top gate voltage V_{g} and drain-to-source potential V_{ds} at the current time moment t_i in the presence of traps, we rely on the previously applied V_{g} at time moment t_{i-1}, on the time difference between the sampling points $t_i - t_{i-1}$, as well as on the traps-induced accumulation ΔV_{traps} at moment t_{i-1}. All these values are then logged in the GNR^2_{up} corresponding table for a wide range of scenarios.

To validate the integrate-and-fire behavior of the proposed graphene-based neuron, we applied as indicated in Figure 3, a deterministic periodic neuronal input V_{in} with $2\,\text{ms}$ spike duration and $5\,\text{ms}$ inter-spike intervals and gradually increased the V_{in} peak amplitude from $100\,\text{mV}$ to $180\,\text{mV}$. We observe that individual V_{in} spikes contributions are gradually accumulated and proportionally reflected in the GNR^2_{up} conductance G increase. Also, we see that for smaller V_{in} spikes ($<180\,\text{mV}$), the conductance increase saturates at a level below the firing threshold, while for $180\,\text{mV}$ V_{in} spikes it can reach the firing threshold, and as a result an output spike event is triggered and reflected in the V_{out} value.

As biochemical processes of individual neurons, as well as surrounding neuronal network activities exhibit stochasticity, the neuronal spike trains exhibit inherent variability. To evaluate the proposed neuron behavior in such conditions, we considered a stochastic input spike train (sampled from a Poisson distribution) with a firing frequency of 50 spikes per second (comparable scenario with that of biological neurons), and added a white Gaussian noise floor with signal-to-noise ratio SNR $= 17$. Figure 4 illustrates the neuron corresponding firing response. We note that every firing event is triggered by an input spike and not by the noise, even though the noise does contribute to the membrane potential accumulation. This suggests that the proposed neuron is robust to input noise. To gain better insight and quantify the variability of the output spike train produced by the proposed neuron, we

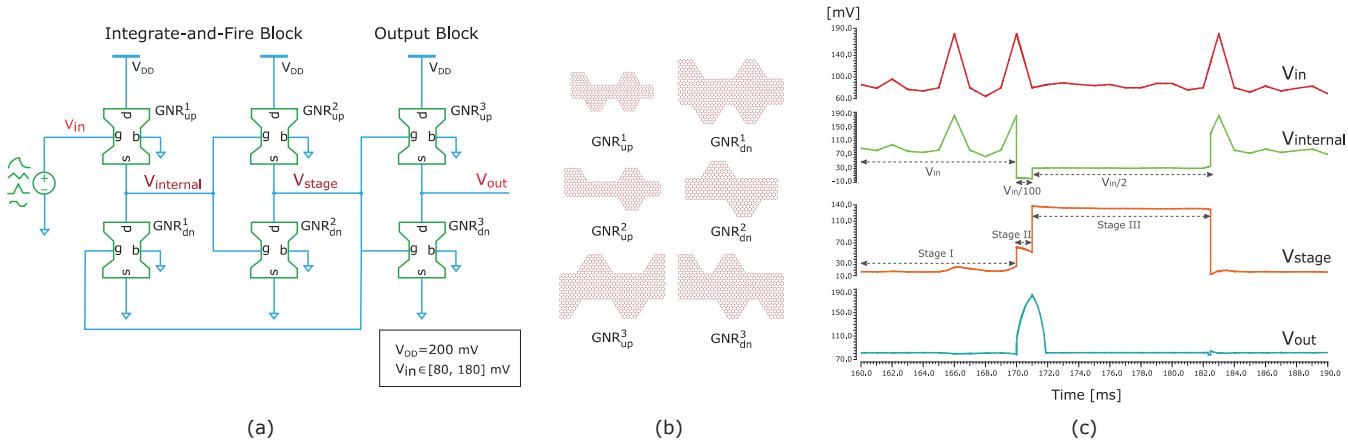

Fig. 2: GNR-based LIF neuron: (a) Neuron structure, (b) GNR topologies, and (c) Basic operation.

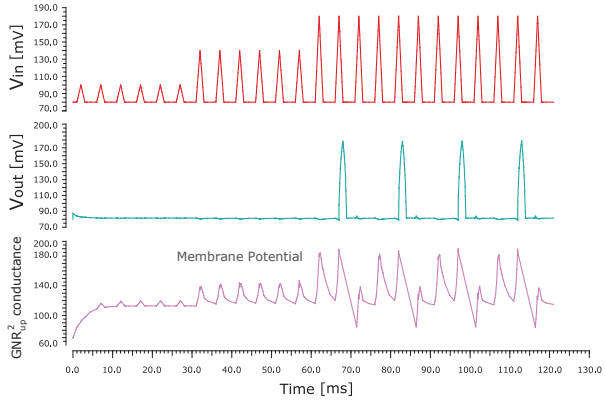

Fig. 3: Integrate and fire dynamics.

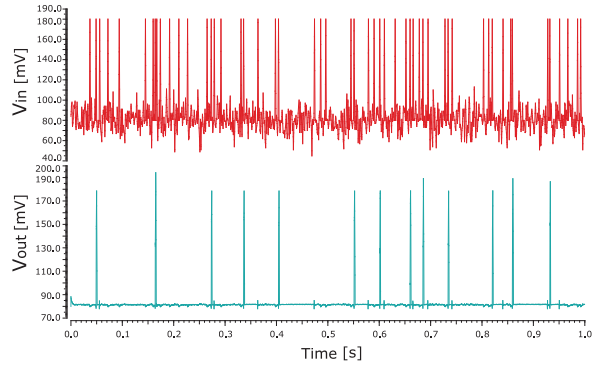

Fig. 4: Graphene-based neuron dynamics under random input.

consider a range of input firing rates from 20 to 200 spikes per second and calculate the output mean firing rate and the variation coefficient CV_{ISI}, which is equal to the standard deviation of the inter-spike timing intervals divided by their mean. Simulation results, depicted in Figure 5, indicate a steady linear increase of the mean output firing rate, suggesting a regular firing behavior for the proposed graphene-based neuron. The output spike train propensity for regularity is also confirmed by a slightly decreasing and < 1 inter-spike interval coefficient of variation.

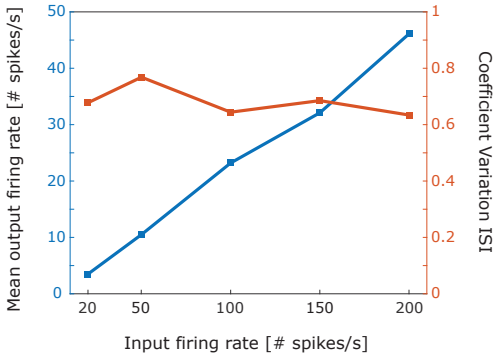

Fig. 5: Output spike statistics for variable input firing rate.

In retrospective, the proposed graphene-based neuron exhibits a small footprint (max. $36\,\mathrm{nm}^2$ per GNR device), and low voltage operation (e.g., $200\,\mathrm{mV}$), which are desired characteristics for artificial neural networks large-scale implementations. Our simulations indicate regularity of firing events under noisy stochastic input spike trains. Furthermore, the considered $2\,\mathrm{ms}$ spike duration and $100\,\mathrm{mV}$ spike amplitude are comparable with that observed in biological neurons, suggesting the potential to fabricate biologically plausible artificial neurons potentially interface-able with biological tissues.

V. Conclusions

In this paper, we proposed an ultra-compact, all graphene-based nonlinear leaky integrate-and-fire neuron. By means of SPICE simulation, we demonstrated that the proposed neuron can properly emulate the basic spiking neuron dynamics under periodic input spikes. We further investigated the output spikes' behavior under stochastic noisy input spike trains. Our simulation results indicated variability resilience and neuronal output firing regularity for a varying input firing rate (from 20 to 200 spikes per second). The small area, low energy (inherent to the $200\,\mathrm{mV}$ supply voltage) and the biologically plausible settings (e.g., $2\,\mathrm{ms}$ and $100\,\mathrm{mV}$ spike duration and amplitude) are certainly enabling factors for the potential implementation of large-scale biocompatible neural systems.

REFERENCES

[1] J. H. B. Wijekoon and P. Dudek, "A CMOS circuit implementation of a spiking neuron with bursting and adaptation on a biological timescale." in *IEEE Biomedical Circuits and Systems Conference (BIOCAS)*, 2009, pp. 193–196.

[2] X. Wu, V. Saxena, K. Zhu and S. Balagopal, "A CMOS spiking neuron for brain-inspired neural networks with resistive synapses and in-situ learning." in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 11, 2015, pp. 1088–1092.

[3] D.B. Strukov, G.S. Snider, D.R. Stewart and R.S. Williams, "The missing memristor found." in *Nature*, vol. 453, no. 7191, 2008, pp. 80–83.

[4] M. Al-Shedivat et al., "Inherently stochastic spiking neurons for probabilistic neural computation." in *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 356–359.

[5] Y. Babacan, F. Kaçar and K. Gürkan, "A spiking and bursting neuron circuit based on memristor." in *Neurocomputing*, vol. 203, 2016, pp. 86–91.

[6] T. Tuma et al., "Stochastic phase-change neurons." in *Nature Nanotechnology*, vol. 11, no. 8, 2016, pp. 693–699.

[7] C.D. Wright, P.Hosseini and D.J.A. Vasquez, "Beyond von-Neumann computing with nanoscale phase-change memory devices." in *Advanced Functional Materials*, vol. 23, no. 18, 2013, pp. 2248–2254.

[8] A. Ferrari et al., "Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems." in *Nanoscale*, vol. 7, no. 11, 2015, pp. 4598–4810.

[9] Y. Y. Chen et al., "Schottky-barrier-type graphene nano-ribbon field-effect transistors: a study on compact modeling, process variation, and circuit performance." in *IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, 2013, pp. 82–88.

[10] X. Yang et al., "Graphene tunneling FET and its applications in low-power circuit design." in *20th Symposium on Great Lakes Symposium on VLSI (GLSVLSI)*, 2010, pp. 263–268.

[11] S. Moysidis, I. G. Karafyllidis, and P. Dimitrakis, "Graphene logic gates." in *IEEE Transactions on Nanotechnology, (TNANO)*, vol. 17, no. 4, 2018, pp. 852–859.

[12] Y. Jiang, N. Cucu Laurenciu, H. Wang, and S.D. Cotofana, "Graphene nanoribbon-based complementary logic gates and circuits," in *IEEE Transactions on Nanotechnology, (TNANO)*, vol. 18, 2019.

[13] H. Wang, N. Cucu Laurenciu, Y. Jiang, and S.D. Cotofana, "Graphene nanoribbon-based synapses with versatile plasticity." in *15th IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, 2019.

[14] ——, "Atomistic-level hysteresis-aware graphene structures electron transport model." in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2019, pp. 1–5.

[15] S. Datta, *Quantum transport: atom to transistor*. Cambridge university press, 2005.

[16] U. Jung et al., "Extraction of the interface state density of top-gate graphene field-effect transistors." in *IEEE Electron Device Letters*, vol. 36, no. 4, 2015, pp. 408–410.

[17] G. Kalon et al., "The role of charge traps in inducing hysteresis: Capacitance–voltage measurements on top gated bilayer graphene." in *Applied Physics Letters*, vol. 99, no. 8, 2011, p. 083109.

[18] Synopsys. [Online]. Available: https://www.synopsys.com/verification/ams-verification/hspice.html/.