

Perceptrons from Memristors

Francisco Silva,^{1,*} Mikel Sanz,^{2,†} João Seixas,^{3,4,‡} Enrique Solano,^{2,5,6,§} and Yasser Omar^{1,7,¶}

¹*Instituto Superior Técnico, Universidade de Lisboa, Portugal*

²*Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, E-48080 Bilbao, Spain*

³*CeFEMA, Instituto Superior Técnico, Universidade de Lisboa, Portugal*

⁴*Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisbon, Portugal*

⁵*IKERBASQUE, Basque Foundation for Science, Maria Diaz de Haro 3, 48013 Bilbao, Spain*

⁶*Department of Physics, Shanghai University, 200444 Shanghai, China*

⁷*Physics of Information and Quantum Technologies Group, Instituto de Telecomunicações, Portugal*

(Dated: July 16, 2018)

Memristors, resistors with memory whose outputs depend on the history of their inputs, have been used with success in neuromorphic architectures, particularly as synapses or non-volatile memories. A neural network based on memristors could show advantages in terms of energy conservation and open up possibilities for other learning systems to be adapted to a memristor-based paradigm, both in the classical and quantum learning realms. No model for such a network has been proposed so far. Therefore, in order to fill this gap, we introduce models for single and multilayer perceptrons based on memristors. We adapt the delta rule to the memristor-based single-layer perceptron and the backpropagation algorithm to the memristor-based multilayer perceptron. We ran simulations of both the models and the training algorithms. These showed that both of them perform well and in accordance with Minsky-Papert's theorem, which motivates the possibility of building memristor-based hardware for a physical neural network.

I. INTRODUCTION

The perceptron, introduced by Rosenblatt in 1958 [1], was one of the first models for supervised learning. In a perceptron, the inputs $x_1 \dots x_n$ are linearly combined with coefficients given by the weights $w_1 \dots w_n$, as well as with a bias b to form the input v to the neuron (see Fig. 1). v is then fed into a non-linear function whose output is either 0 or 1. The goal of the perceptron is thus to find a set of weights $\{w_i\}$ that correctly assigns inputs $\{x_i\}$ to one of two predetermined binary classes. The correct weights for this task are found by an iterative training process, for instance the delta rule [2]. However, the perceptron is only capable of learning linearly separable patterns, as was shown in 1969 by Minsky and Papert [3]. These limi-

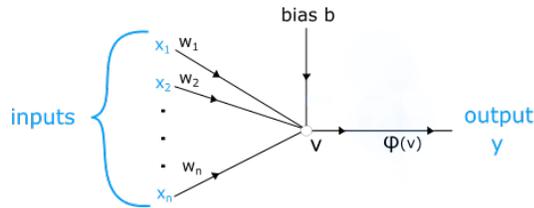


Fig. 1. In a single-layer perceptron (SLP) the inputs x_i are multiplied by their respective weights w_i and added, together with a bias b to form the net input to the SLP, v . The output y of the SLP is given by some activation function, $\phi(v)$.

tations triggered a search for more capable models, which eventually resulted in the proposal of the multilayer perceptron. These objects can be seen as several layers of perceptrons connected to each other by synapses (see Fig. 2). This structure ensures that the multilayer perceptron does not suffer from the same limitations as Rosenblatt's perceptron. In fact, a multilayer perceptron with at least one hidden layer of neurons and with conveniently chosen activation functions can approximate any continuous function to an arbitrary accuracy [4].

There are various methods to train a neural network such as the multilayer perceptron. One of the most widespread is the backpropagation algorithm, a generalization of the original delta rule [5]. Artificial neural

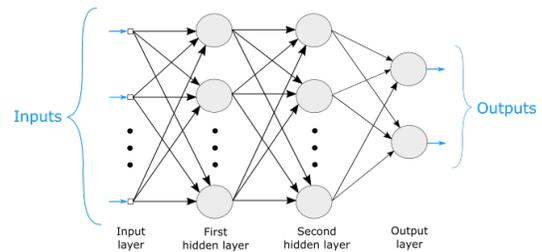


Fig. 2. In a multilayer perceptron (MLP), single-layer perceptrons (SLP) are arranged in layers and connected to each other, with the outputs of the SLPs in the output layer being the outputs of the MLP. Here, each SLP is represented by a disc.

networks such as the multilayer perceptron have proven extremely useful in solving a wide variety of problems [6–8], but they have thus far mostly been implemented in digital computers. This means that we are not profiting from some of the advantages that these networks could have over traditional computing paradigms, such as very

* francisco.horta.ferreira.da.silva@tecnico.ulisboa.pt

† mikel.sanz@ehu.eus

‡ joao.seixas@tecnico.ulisboa.pt

§ enr.solano@gmail.com

¶ yasser.omar@lx.it.pt

low energy consumption and massive parallelization [9]. Keeping these advantages is, of course, of utmost interest, and this could be done if a physical neural network was used instead of a simulation on a digital computer. In order to construct such a network, a suitable building block must be found, with the memristor being a good candidate.

The memristor was first introduced in 1971 as a two-terminal device that behaves as a resistor with memory [10]. The three known elementary circuit elements, namely the resistor, the capacitor and the inductor, can be defined by the relation they establish between two of the four fundamental circuit variables: the current i , the voltage u , the charge q and the flux-linkage ϕ . There are six possible combinations of these four variables, five of which lead to widely-known relations: three from the circuit elements mentioned above, and two given by $q(t) = \int_{-\infty}^t i(\tau)d\tau$ and $\phi(t) = \int_{-\infty}^t u(\tau)d\tau$. This means that only the relation between ϕ and q remains to be defined: the memristor provides this missing relation. Despite having been predicted in 1971 using this argument, it was not until 2008 that the existence of memristors was demonstrated at HP Labs [11], which led to a new boom in memristor-related research [12]. In particular, there have been proposals of how memristors could be used in Hebbian learning systems [13–15], in the simulation of fluid-like integro-differential equations [16], in the construction of digital quantum computers [17] and of how they could be used to implement non-volatile memories [18].

The pinched current-voltage hysteresis loop inherent to memristors endows them with intrinsic memory capabilities, leading to the belief that they might be used as a building block in neural computing architectures [19–21]. It is then natural to ask if they can be used as a basic unit in artificial neural networks. The main goal of this paper is to introduce a memristor-based perceptron, i.e., a single-layer perceptron (SLP) built from memristors, which will be generalized to a memristor-based multi-layer perceptron (MLP). We will also introduce learning rules for both perceptrons, based on the delta rule for the SLP, and on the backpropagation algorithm for the MLP. It is important to note that while there have been proposals of how memristors can be used in learning systems [13–15], these were only regarding Hebbian learning, which constitutes a different learning paradigm than the perceptrons we are proposing.

II. THE MEMRISTOR AS A DYNAMICAL SYSTEM

In general, a current-controlled memristor is a dynamical system whose evolution is described by the following pair of equations [10]

$$\begin{cases} V = R(\vec{\gamma}, I)I, & (1a) \\ \dot{\vec{\gamma}} = \vec{f}(\vec{\gamma}, I). & (1b) \end{cases}$$

The first one is Ohm's law and relates the voltage output of the memristor V with the current input I through the memristance $R(\vec{\gamma}, I)$, which is a scalar function depending both on I and on the set of the memristor's internal variables $\vec{\gamma}$. This dependence of the memristance on the internal variables induces the memristor's output dependence on past inputs, i.e., this is the mechanism that endows the memristor with memory. The second equation describes the time-evolution of the memristor's internal variables by relating their time derivative, $\dot{\vec{\gamma}}$, to an n -dimensional vector function $\vec{f}(\vec{\gamma}, I)$, depending on both previous values of the internal variables and the input of the memristor.

A. Memristor-based Single-Layer Perceptron

Our goal is to implement a perceptron and an adaptation of the delta rule to train it using only a memristor. To this end, we use the memristor's internal variables to store the SLP's weights. Equation (1b) allows us to control the evolution of the memristor's internal variables and implement a learning rule. If, for example, we want to implement a SLP with two inputs we need a memristor with three internal variables, two of them to store the weights of the connections between the inputs and the SLP and the other one to store the SLP's bias weight.

Let us then consider a memristor with three internal state variables, from now on labeled by $\vec{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ and in which $\vec{f} = (f_1, f_2, f_3)$. It could be difficult to externally control multiple internal variables. However, a possible solution is to several memristors with the chosen requirements and with an externally controlled internal variable each.

In order to understand the form of these functions, we must remember that we expect different behaviours from the perceptron depending on the stage of the algorithm. In the forward propagation stage, the weights must remain constant to obtain the output for a given input. In this phase the internal variables must not change. On the other hand, in the backpropagation stage, we want to update the perceptron's weights by changing the internal variables. However, it may happen that the update is different for each of the weights, so we need to be able to change only one of the internal variables without affecting the others.

There are thus three different possible scenarios in the backpropagation stage: we want to update γ_1 , while γ_2 and γ_3 should not change; we want to update γ_2 , while γ_1 and γ_3 should not change, and we want to update γ_3 , while γ_1 and γ_2 should not change. To conciliate this with the fact that a memristor takes only one input, we propose the use of threshold-based functions, as well as a bias current I_b , for the evolution of the internal variables

$$V(t) = g(I, \gamma_1, \gamma_2, \gamma_3), \quad (2)$$

$$\dot{\gamma}_i = (I - I_b)\theta(I - I_{\gamma_i}), \quad (3)$$

where g is an activation function and θ is the Heaviside function. If I_{γ_1} , I_{γ_2} and I_{γ_3} are sufficiently different from each other and from zero, we can reach the correct behaviour by choosing the memristor's input appropriately. In the aforementioned construction in which our memristor with three internal variables is constructed as an equivalent memristor, we can also use an external current or voltage control to keep the internal variable fixed. In fact, this is how it is usually addressed experimentally [21–24]. Therefore, we can assume that this construction is possible. It is important to note that, in an experimental implementation, this threshold system does not need to be based on the input currents' intensities. It can, for instance, be based on the use of signals of different frequencies for each of the internal variables or in the codification of the signals meant for each of the internal variables in AC voltage signals.

Algorithm 1 Delta rule for Single-layer Perceptron

INITIALIZATION

Set the bias current I_b to 0.

Initialize the weights w_1 , w_2 , w_b .

Set the internal state variables γ_1 , γ_2 , γ_3 to w_1 , w_2 and w_b , respectively.

for d in data **do**

FORWARD PASS

Compute the net input to the perceptron:

$$I = w_1x_1 + w_2x_2. \quad (4)$$

Compute the perceptron's output:

$$V = g(I, \gamma_1, \gamma_2, \gamma_3). \quad (5)$$

BACKWARD PASS

Compute the difference Δ between the target output and the actual output:

$$\Delta = T - V. \quad (6)$$

Compute the derivative of the activation function with respect to the net input, g' .

for i in internal variables **do**

Set the bias $I_b = I_{\gamma_i}$.

Update γ_i by inputting $I = \Delta x_i g' + I_b$.

end for

Update the weights by setting them to the updated values of the internal state variables.

Set the bias $I_b = 0$.

end for

We are now ready to present a learning algorithm for our SLP based on the delta rule, which is described in Algorithm 1. In case one wants to generalize this procedure to an arbitrary number of inputs n , this can be trivially achieved by using a memristor with $n+1$ internal variables and adapting Algorithm 1 accordingly.

B. Memristor-based Multilayer Perceptron

In this model, memristors are used to emulate both the connections and the nodes of a MLP. In principle, the nodes could be emulated by non-linear resistors, but using memristors allows us to take advantage of their internal variable to implement a bias weight, which in some cases proves fundamental for a successful network training.

The equations describing the evolution of the memristor at each node in this model are the same as in the seminal HP Labs paper [11]. We have chosen the experimentally tested set

$$V(t) = \left(R_{\text{ON}} \frac{\gamma(t)}{D} + R_{\text{OFF}} \left(1 - \frac{\gamma(t)}{D} \right) \right) I(t), \quad (7)$$

$$\dot{\gamma} = \begin{cases} \mu_V \frac{R_{\text{ON}}}{D} I(t) - I_\gamma & \text{if } \mu_V \frac{R_{\text{ON}}}{D} I(t) > I_\gamma, \\ 0 & \text{o.w.} \end{cases} \quad (8)$$

Here, R_{ON} and R_{OFF} are, respectively, the doped and undoped resistances of the memristor, D and μ_V are physical memristor parameters, namely the thickness of its semiconductor film and its average ion mobility, and I_γ is a threshold current playing the same role as the $I_{\bar{\gamma}}$ in the model for the memristor-based SLP introduced above. Equation (7) can be approximated by

$$V(t) = R_{\text{OFF}} \left(1 - \frac{\gamma(t)}{D} \right) I(t), \quad (9)$$

since we have that $\frac{R_{\text{ON}}}{R_{\text{OFF}}} \approx \frac{1}{100}$. If, for instance, we impose a constant current input I to the memristor for a time t , the output is given by

$$V(t) \propto -I^2 t. \quad (10)$$

This can be achieved in practice by using a current integrator. It is then possible to implement non-linear activation functions starting from Equation (7), which is an important condition for the universality of neural networks [25].

Looking now at synaptic memristors, their evolution is described by

$$V(t) = \gamma(t)I(t), \quad (11)$$

$$\dot{\gamma} = \left(\mu_V \frac{R_{\text{ON}}}{D} I(t) - I_\gamma \right) \theta \left(\mu_V \frac{R_{\text{ON}}}{D} I(t) - I_\gamma \right). \quad (12)$$

In synaptic memristors, the internal variable γ is used to store the weight of the respective connection, whereas in node memristors the internal variable is used to store the node's bias weight.

As explained before, the node memristors are chosen to operate in a non-linear regime, which allows us to implement non-linear activation functions. On the other hand, we choose a linear regime for synaptic memristors,

which allows us to emulate the multiplication of weights by signals.

It must be mentioned that Equation (8) is only valid $\gamma \in [0, D]$. If we were to store the network weights in the internal variables using only a rescaling constant A , i.e., $w = A\gamma$, then the weights would all have the same sign. Although convergence of the standard backpropagation algorithm is still possible in this case [26], it is usually slower and more difficult, so it is convenient to redefine the variable [11] $D \rightarrow D'$ so that the interval of the internal variable in which Equation (8) is valid becomes $[-D'/2, D'/2]$. Using a rescaling constant B , the network weights can then be in the interval $[-BD'/2, BD'/2]$.

The new learning algorithm is an adaptation of the backpropagation algorithm, chosen due to its widespread use and robustness. In our case, the activation function of the neurons is the function that relates the output of a node memristor with its input, as seen in Equation (7). The local gradients of the output layer and hidden layer neurons respectively given by:

$$\left\{ \begin{array}{l} \text{Output: } \delta_k = T_k \phi' \left(\sum_i V_{ik} \right), \\ \text{Hidden: } \delta_k = \phi' \left(\sum_i V_{ik} \right) \sum_l \delta_l w_{kl}. \end{array} \right. \quad (13)$$

$$\left\{ \begin{array}{l} \text{Output: } \delta_k = T_k \phi' \left(\sum_i V_{ik} \right), \\ \text{Hidden: } \delta_k = \phi' \left(\sum_i V_{ik} \right) \sum_l \delta_l w_{kl}. \end{array} \right. \quad (14)$$

In Equation (13), T_k denotes the target output for neuron k in the output layer. In Equations (13) and (14), ϕ' is the derivative of the neuron's activation function with respect to the input to the neuron $\sum_i V_{ik}$. Finally, in Equation (14), the sum $\sum_l \delta_l w_{kl}$ is taken over the gradients of all neurons l in the layer to the right of the neuron that are connected to it by weights w_{kl} . The update to the bias weight of a node memristor is given by:

$$\Delta w_k = \eta \delta_k, \quad (15)$$

where η is the learning rate. The connection weight w_{ij} is updated using $\Delta w_{ij} = \eta \delta_j V_i$, where δ_j is the local gradient of the neuron to the right of the connection, and V_i is the output of the neuron to the left of the connection.

We count now with all necessary elements to adapt the backpropagation algorithm for our memristor-based MLP, as described in Algorithm 2.

Algorithm 2 Backpropagation for Multilayer Perceptron

INITIALIZATION

Set the bias current I_b to 0.

Initialize the weights $\{w_{ij}\}$ and $\{w_{b_k}\}$.

Set the internal variable γ_{ij} of each connection memristor ij to the respective connection weight w_{ij} .

Set the internal variable γ_k of each connection memristor k to the respective bias weight w_{b_k} .

for d in data **do**

FORWARD PASS

for l in layers **do**

Compute the output of each connection memristor ij in layer l :

$$V_{ij}(w_{ij}, I) = w_{ij}I. \quad (16)$$

Sum the outputs of the connection memristors connected to each node memristor k in layer l

$$\text{in}_k = \sum I_{ik} \quad (17)$$

Compute the node memristor's output:

$$V_k = R_{\text{OFF}} \left(1 - \frac{\gamma_{b_k}}{D} + \frac{R_{\text{ON}}}{R_{\text{OFF}}} \frac{\gamma_{b_k}}{D} \right) \text{in}_k.$$

end for

BACKWARD PASS

for k in output layer **do**

Compute the difference Δ between the target output and the actual output of the node memristor:

$$\Delta_k = T_k - V_k. \quad (18)$$

Compute the local gradient of the node memristor using Equation (13).

end for

for layer in hidden layers **do**

for node in layer **do**

Compute the local gradient of node memristor l in layer using Equation (14).

end for

end for

for connection in connections **do**

Compute the weight update.

Set the bias current: $I_b = I_{\gamma_{ij}}$.

Update the connection memristor's internal variable by inputting $I = \Delta w_{ij} + I_b$ to it.

Update the connection's weight by setting it to the updated value of the respective internal variable.

end for

for node in nodes **do**

Compute the bias weight update according to Equation (15).

Set the bias current: $I_b = I_{\gamma_b}$.

Update the node memristor's internal variable by inputting $I = \Delta w_k + I_b$.

Update the bias weight by setting it to the updated value of the respective internal variable.

end for

end for

III. SIMULATION RESULTS

In order to test the validity of our SLP and MLP, we tested their performance on three logical gates: OR, AND and XOR. The first two are simple problems which should be successfully learnt by SLP and MLP, whereas only the MLP should be able to learn the XOR gate, due to Minsky-Papert's theorem.

The Glorot weight initialization scheme [27] was used for all simulations, as it has been shown to bring faster convergence in some problems when compared to other initialization schemes. In this scheme the weights are initialized according to $\mathcal{U}(-1, 1)$, weighed by $\sqrt{\frac{6}{n_{in}+n_{out}}}$, where n_{in} and n_{out} are the number of neurons in the previous and following layers, respectively. The data sets used contain 100 randomly generated labeled elements, which were shuffled for each epoch, and the cost function is:

$$E = \frac{1}{2}(T - O)^2, \quad (19)$$

where T is the target output and O the actual output.

A. Single-Layer Perceptron Simulation Results

For the SLP, a learning rate of 0.1 was used for all tested gates, a value set by trial and error. The metric we used to evaluate the evolution of the network's performance on a given problem was its total error over an epoch, which is given by Equation (20).

$$E_{\text{total}} = \sum_j E_j = \frac{1}{2} \sum_j (T_j - O_j)^2, \quad (20)$$

where the sum is taken over all elements in the training set. In Fig. 3, the evolution of the total error over 1000 epochs, averaged over 100 different realizations of the starting weights, is plotted.

We observe that our SLP successfully learns the gates OR and AND, with the total error falling to 0 within 200 epochs, as expected from a SLP. However, the total error of our SLP for the XOR gate does not go to zero, which means that it is not able to learn this gate, in accordance with Minsky-Papert's theorem.

B. Multilayer Perceptron Simulation Results

The structure of the network was chosen following [28]. There, a network with one hidden layer of two neurons is recommended for the case of two inputs and one output. As noted in [28], networks with only one hidden layer are capable of approximating any function, although in some problems, adding extra hidden layers improves the performance. However, the results obtained by

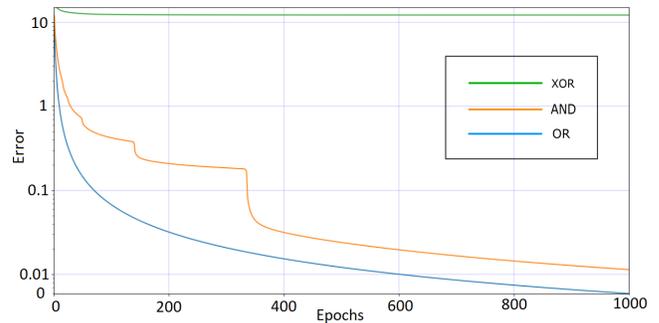


Fig. 3. Evolution of the learning progress of our single-layer perceptron (SLP), quantified by its total error, given by Equation (20), for the OR, AND and XOR gates over 1000 epochs. The total error of our SLP for the OR and AND gates goes to 0 very quickly, indicating that our SLP successfully learns these gates. The same is not true for the XOR gate, which our SLP is incapable of learning, in accordance with Minsky-Papert's theorem [3].

employing only one hidden layer are satisfactory, thus there is no need for a more complex network structure. There is also the matter of how many neurons must be employed in the hidden layer. In this case, there is a trade-off between speed of training and accuracy. A network with more neurons in the hidden layer counts with more free parameters, so it will be able to output a more accurate fit, but at the cost of a longer time required to train the network. A rule of thumb for choosing the number of neurons in the hidden layer is to start with an amount that is between the number of inputs and the number of outputs and adjust according to the results obtained. This leads to two neurons for the hidden layer and, similarly to what happened with the number of hidden layers, the results obtained using two neurons in the hidden layer are sufficiently accurate, so there was no need to try other structures. The learning rates used, which we have chosen through trial and error, are 0.1 for the OR and AND gates, and 0.01 for the XOR gate. In Fig. 4, the evolution of the total error over 1000 epochs, averaged over 100 different realizations of the starting weights, is plotted.

As was the case for our SLP, our MLP successfully learns the OR and AND gates. In fact, it is able to learn them faster than our SLP, which is a consequence of the larger number of free parameters. Additionally, it is able to learn the XOR gate, indicating that it behaves as well as a regular MLP.

In summary, both memristor-based perceptrons behave as expected. Our SLP is able to learn the OR and AND gates, but not the XOR gate, so it is limited to solving linearly separable problems, just as any other single-layer neural network. However, our MLP is not subject to such a limitation and it is able to learn all three gates.

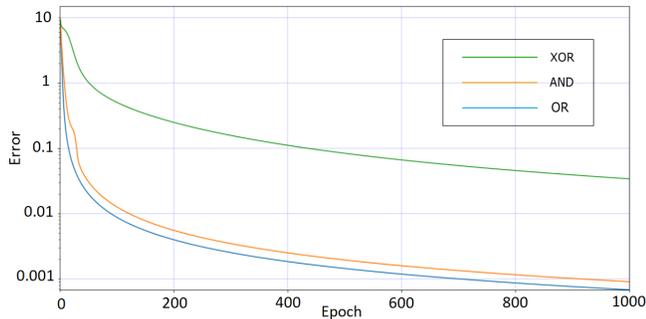


Fig. 4. Evolution of the learning progress of our multilayer perceptron (MLP), quantified by its total error, given by Equation (20) for the OR, AND and XOR gates over 1000 epochs. As can be seen, the total error of our MLP for these gates approaches 0, indicating that it successfully learns all three gates.

C. Receiver Operating Characteristic Curves

As another measure of the perceptrons' performance, we show in Fig. 5 the receiver operating characteristic (ROC) curves obtained with perceptrons trained for 500 epochs on data sets of size 100. The curves shown were obtained using a SLP trained for the OR gate, a SLP trained for the XOR gate and a MLP trained for the XOR gate, with thresholds of $t = 0.3, 0.5$ and 0.7 for each. Again, we see that the SLP is capable of learning the OR gate but not XOR, since it correctly classifies the inputs for OR every time, but its performance is equivalent to random guessing for XOR. We can also see that the MLP is capable of learning the XOR gate, since it correctly classifies its inputs every time. The learning rates used in training were 0.1 for the SLP on both gates and 0.01 for the MLP on XOR gate, as explained in the previous subsection.

IV. CONCLUSION

In this paper, we introduced a model for single and multilayer perceptrons based on memristors. We provide learning algorithms for both, based on the delta rule and on the backpropagation algorithm, respectively. Using a threshold-based system, our model is able to use the internal variables of memristors to store and update the perceptron's weights. We also ran simulations of both models, which revealed that they behaved as expected and in accordance with Minsky-Papert's theorem. Our memristor-based perceptrons have the same capabilities

and are subject to the same limitations of regular perceptrons and show the feasibility and power of a memristor-based neural network.

Our models are the first ones in which memristors are used as both the nodes and the synapses, thus paving

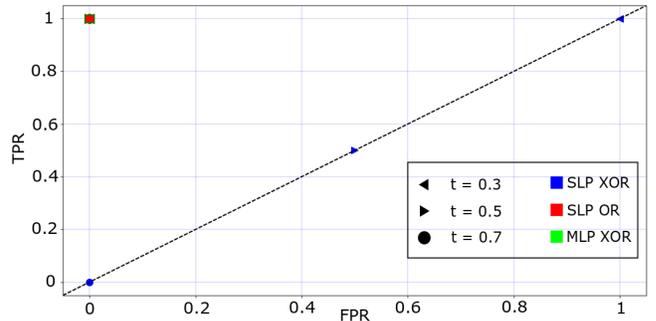


Fig. 5. ROC curves obtained with the SLP for the OR and XOR gates, and with the MLP for the XOR gate. The thresholds used were $t = 0.3, 0.5$ and 0.7 . We can see that the SLP correctly classifies the inputs for the OR gate every time, but it does not perform better than random guessing for the XOR gate, as expected. On the other hand, the MLP correctly classifies the XOR gate inputs every time.

the way for other neural network architectures and algorithms based on memristors. In particular, it would be interesting to try to extend these models to the quantum computing paradigm, using a recently proposed quantum memristor [29], and its implementation in different quantum technologies, such as superconducting circuits [30] or quantum photonics [31].

ACKNOWLEDGMENTS

Work by FS was supported in part by a New Talents in Quantum Technologies scholarship from the Calouste Gulbenkian Foundation. MS and ES are grateful for the funding of Spanish MINECO/FEDER FIS2015-69983-P and Basque Government IT986-16. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), under field work proposal number ERKJ335. YO thanks support from Fundação para a Ciência e a Tecnologia (Portugal), namely through programmes PTDC/POPH/POCH and projects UID/EEA/50008/2013, IT/QuNet, partially funded by EU FEDER, from the QuantERA project TheBlinQC, and from the JTF project NQuN (ID 60478).

[1] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain."

Psychological review, vol. 65, no. 6, p. 386, 1958.

- [2] B. Widrow and M. E. Hoff, "Adaptive switching circuits," Stanford Univ Ca Stanford Electronics Labs, Tech. Rep., 1960.
- [3] M. Minsky, S. A. Papert, and L. Bottou, *Perceptrons: An introduction to computational geometry*. MIT press, 2017.
- [4] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, p. 533, 1986.
- [6] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 1, pp. 23–38, 1998.
- [7] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz, and J. Makhoul, "Fast and robust neural network joint models for statistical machine translation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2014, pp. 1370–1380.
- [8] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Transactions on biomedical engineering*, vol. 41, no. 9, pp. 837–845, 1994.
- [9] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.
- [10] L. Chua, "Memristor-the missing circuit element," *IEEE Transactions on circuit theory*, vol. 18, no. 5, pp. 507–519, 1971.
- [11] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, p. 80, 2008.
- [12] T. Prodromakis and C. Toumazou, "A review on memristive devices and applications," in *Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 934–937.
- [13] D. Soudry, D. Di Castro, A. Gal, A. Kolodny, and S. Kvatinsky, "Hebbian learning rules with memristors," *Israel Institute of Technology: Haifa, Israel*, 2013.
- [14] K. D. Cantley, A. Subramaniam, H. J. Stiegler, R. A. Chapman, E. M. Vogel *et al.*, "Hebbian learning in spiking neural networks with nanocrystalline silicon tfts and memristive synapses," *IEEE Transactions on Nanotechnology*, vol. 10, no. 5, pp. 1066–1073, 2011.
- [15] W. He, K. Huang, N. Ning, K. Ramanathan, G. Li, Y. Jiang, J. Sze, L. Shi, R. Zhao, and J. Pei, "Enabling an integrated rate-temporal learning scheme on memristor," *Scientific reports*, vol. 4, p. 4755, 2014.
- [16] G. A. Barrios, J. Retamal, E. Solano, and M. Sanz, "Analog simulator of integro-differential equations with classical memristors," *arXiv preprint arXiv:1803.05945*, 2018.
- [17] Y. V. Pershin and M. Di Ventra, "Neuromorphic, digital, and quantum computation with memory circuit elements," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 2071–2080, 2012.
- [18] Y. Ho, G. M. Huang, and P. Li, "Nonvolatile memristor memory: device characteristics and design implications," in *Computer-Aided Design-Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*. IEEE, 2009, pp. 485–490.
- [19] F. L. Traversa and M. Di Ventra, "Universal memcomputing machines," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 11, pp. 2702–2715, 2015.
- [20] Y. V. Pershin and M. Di Ventra, "Experimental demonstration of associative memory with memristive neural networks," *Neural Networks*, vol. 23, no. 7, pp. 881–886, 2010.
- [21] J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive devices for computing," *Nature nanotechnology*, vol. 8, no. 1, p. 13, 2013.
- [22] Q. Xia, W. Robinett, M. W. Cumbie, N. Banerjee, T. J. Cardinali, J. J. Yang, W. Wu, X. Li, W. M. Tong, D. B. Strukov *et al.*, "Memristor-cmos hybrid integrated circuits for reconfigurable logic," *Nano letters*, vol. 9, no. 10, pp. 3640–3645, 2009.
- [23] D. Yu, H. H.-C. Iu, Y. Liang, T. Fernando, and L. O. Chua, "Dynamic behavior of coupled memristor circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 6, pp. 1607–1616, 2015.
- [24] R. K. Budhathoki, M. P. Sah, S. P. Adhikari, H. Kim, and L. Chua, "Composite behavior of multiple memristor circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, pp. 2688–2700, 2013.
- [25] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.
- [26] F. Dickey and J. DeLaurentis, "Optical neural networks with unipolar weights," *Optics communications*, vol. 101, no. 5-6, pp. 303–305, 1993.
- [27] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.
- [28] S. Walczak and N. Cerpa, "Heuristic principles for the design of artificial neural networks," *Information and software technology*, vol. 41, no. 2, pp. 107–117, 1999.
- [29] P. Pfeiffer, I. Egusquiza, M. Di Ventra, M. Sanz, and E. Solano, "Quantum memristors," *Scientific reports*, vol. 6, p. 29507, 2016.
- [30] J. Salmilehto, F. Deppe, M. Di Ventra, M. Sanz, and E. Solano, "Quantum memristors with superconducting circuits," *Scientific reports*, vol. 7, p. 42044, 2017.
- [31] M. Sanz, L. Lamata, and E. Solano, "Invited article: Quantum memristors in quantum photonics," *APL Photonics*, vol. 3, no. 8, p. 080801, 2018.