Neural network based analysis of Random Telegraph Noise in Resistive Random Access Memories

Authors: G. González-Cordero[1], M.B. González[2], A. Morell[3], F. Jiménez-Molinos[1], F. Campabadal[2], J.B. Roldán[1]

[1]Departamento de Electrónica y Tecnología de Computadores. Universidad de Granada. Facultad de Ciencias. Avd. Fuentenueva s/n, 18071 Granada, Spain.

[2]Institut de Microelectrònica de Barcelona, IMB-CNM (CSIC), carrer dels Til·lers, S/N. Campus UAB, 08193 Bellaterra, Spain

[3]Departament de Telecomunicació i Enginyeria de Sistemes. Universitat Autònoma de Barcelona. Escola d'Enginyeira, carrer de les Sitges, S/N, 08193 Bellaterra, Spain

Electronic mail: *jroldan@ugr.es*

Abstract

The characterization of Random Telegraph Noise (RTN) signals in Resistive Random Access Memories (RRAM) is a challenge. The inherent stochastic operation of these devices, much different to what is seen in other electron devices such as MOSFETs, diodes, etc., makes this issue more complicated from the mathematical viewpoint. Nevertheless, the accurate modeling of these type of signals is essential for their use in digital and analog applications. RTN signals are revealed to be linked to the emission and capture of electrons by traps close to the conductive filament (CF) that can influence resistive switching (RS) operation in RRAMs. RTN features depend on the number of active traps, on the interaction between these traps at different times, on the occurrence of anomalous effects, etc. Using a new representation technique, the Locally Weighted Time Lag Plot (LWTLP), a highly efficient method in terms of computation, data from current-time (I-t) traces can be represented with a pattern that allows the analysis of important RTN signal features. In addition, Self-organizing maps (SOM), a neural network devoted to clustering, can be employed to perform an automatic classification of the RTN traces that have similar LWTLP patterns. This pattern analysis allows a better understanding of RTN signals and the physics underlying them. The new technique presented can be performed in a reasonable computing time and it is particularly adequate for long (I-t) traces. We introduce here this technique and the most important results that can be drawn when applied to long RTN traces experimentally obtained in RRAMs.

Keywords: resistive memories, RRAM, RTN, LWTLP, neural network, self organizing maps, random telegraph noise

1.- INTRODUCTION

New emerging non-volatile memory technologies are getting wide attention among electronics manufacturers and the academic community. These new devices show a set of advantages that makes the viability of high-performance and cost-effective applications a reasonable possibility [1-4]. The applications are not only linked to non-volatile storage-class memory modules but also to neuromorphic circuits and hardware security implementations [2, 5-7]. The growing prominence of mobile devices and the popularity of Internet-of-things (IoT) is pushing the technology towards the use of devices such as RRAMs that can be easily scaled and show fast read and writing times, low power operation, non-volatility and CMOS compatibility.

A resistive memory is made of metal or highly doped semiconductor electrodes and a dielectric in between the electrodes that can be fabricated by different types of oxides and even 2D materials. A wide variety of materials has been employed for dielectric and electrodes [1, 2]. These devices have been successfully used at the integrated circuit level with a promising degree of scalability [8-10]. However, prior to massive industrial use, key issues such as variability and reliability have to be improved. In addition, characterization and modeling have to be pushed forward. Although great efforts from the physical simulation [3, 11-17] and compact modeling [18-22] have been made, new studies in terms of characterization and modeling are needed. With respect to modeling, noise is a significant issue to deal with, in particular RTN, a noise source that is important in RRAMs [5, 23-27].

The CF creation and rupture gives rise to resistive switching (RS) operation. The presence of single or multiple traps inside or close to the CFs influences charge conduction and produces current fluctuations that can lead to RTN [23-25]. These fluctuations can reduce noise margin in memory cell arrays [24, 25] or affect the analog behavior of the devices necessary to mimic electronic synapses in neuromorphic circuits [28], posing important hurdles to the use of these devices in highly-scaled integrated circuits. RTN fluctuations can also be beneficial, for instance, when used as entropy sources in random number generators [23, 29, 30].

Several methods to characterize RTN can be found in the literature [24, 25, 27, 31- 35]. Different facets of RTN signals can be studied making use of them; however, if long (I-t) traces are considered some of these procedures are not appropriate due to their heavy burden from the computational viewpoint. We have recently presented a methodology based on the Locally Weighted Time Lag Plot (LWTLP) [25]. This new method allows a quick analysis of (I-t) RTN signals.

Making use of an experimental (I-t) trace as a starting point, a pattern based on the LWTLP is obtained that produces information about the current levels, their frequency, the transitions between different current levels, etc [25, 32, 36]. In a long trace, different characteristic patterns can be obtained if the trace is divided in regular time intervals, since the traps that influence current fluctuations get activated or deactivated as the measurement goes on. In this respect, the number of active traps, i.e., the different RTN current levels, evolve with time. If a long (I-t) trace with millions of measurements is divided in time windows, we would obtain a characteristic pattern for each of these windows. The question that reasonably comes up is connected with the possibility of repetition of certain patterns, or if any pattern could be characterized by a finite set of patterns. In implementing this analysis, we enter in the data mining/machine learning realm and in particular we deal with a (unsupervised) clustering problem. We have adopted in this work an Artificial Neural Network (ANN) approach using Self-Organizing Maps (SOM), also known as Kohonen networks [37, 38]. SOM has been employed instead of k-Means approach [39] due to the topology preservation and 2-D visualization (as a map) properties.

The paper scheme is the following, section II describes the device fabrication and measurement procedures and section III is devoted to a summary of some of the techniques to analyze RTN signals. In section IV, the new technique is explained and in section V the main results are discussed. Finally, the conclusions are drawn in Sec. VI.

2.- DEVICE FABRICATION AND MEASUREMENT

The fabricated devices were based on the Ni/HfO_2/Si stack, with n-type silicon with a resistivity of (7-13) mΩ·cm. A 20nm-thick HfO_2 layer was deposited by ALD at 498K

using TDMAH and H_2O as precursors. The structures are 5x5μm^2 square cells. The cross-sectional view of the device is shown in the inset of Figure 1. More detailed information about the fabrication process flow is given in Refs. [40, 41].

The current was recorded in the time domain by means of a HP-4155B semiconductor parameter analyzer. The voltage was applied to the top Ni electrode, while the Si substrate was grounded. Both, the I-V curves for the forming process and RS cycles were measured making use of ramped voltage signals (some set/reset curves are shown in Figure 1). Two clearly differentiated resistive states were detected (see Figure 1). The low resistance state (LRS) shows up when the CF is fully formed and the two electrodes are shorted; if the CF is ruptured in a reset process, the device conductance drops off resulting in the high resistance state (HRS).

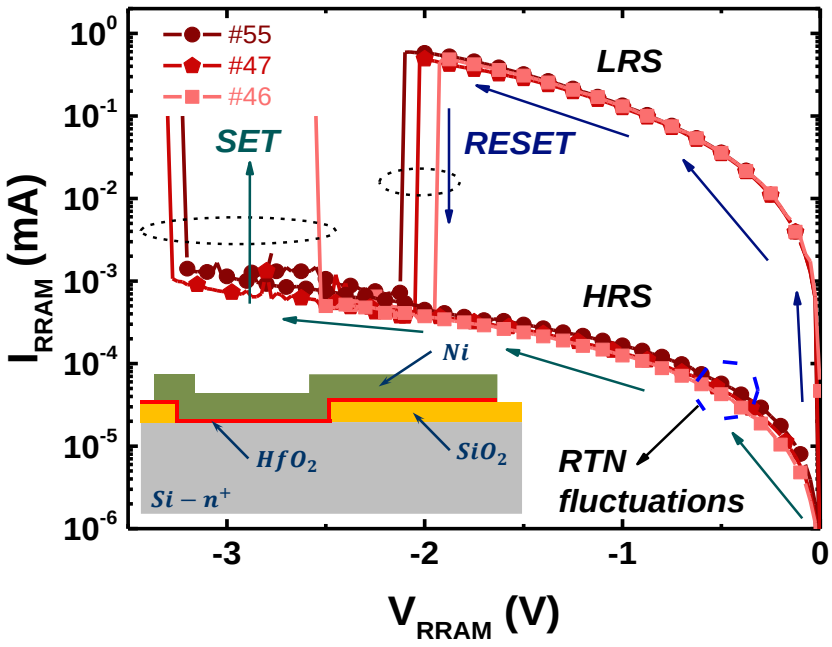


Figure 1. Current versus applied voltage for different set and reset cycles (#55, #46 and #47) within a long RS series of more than a thousand curves, I_{CC}=100μA. Inset: device schematic cross section. The device operation region where RTN signals are measured is highlighted.

RTN data were measured in the HRS at a constant bias of V_{RRAM}=-0.5V for time intervals of several seconds (thousands of data were measured). The RTN signals were recorded by an automatic algorithm that considers the previous measured data making use of a smart procedure (a Matlab® software tool was employed to control the

instrumentation) [41, 36]. Figure 2 shows a long register of approximately 1.2 million samples obtained for 10500 seconds and the details of two time intervals of 10 seconds (where typical RTN signals can be recognized).

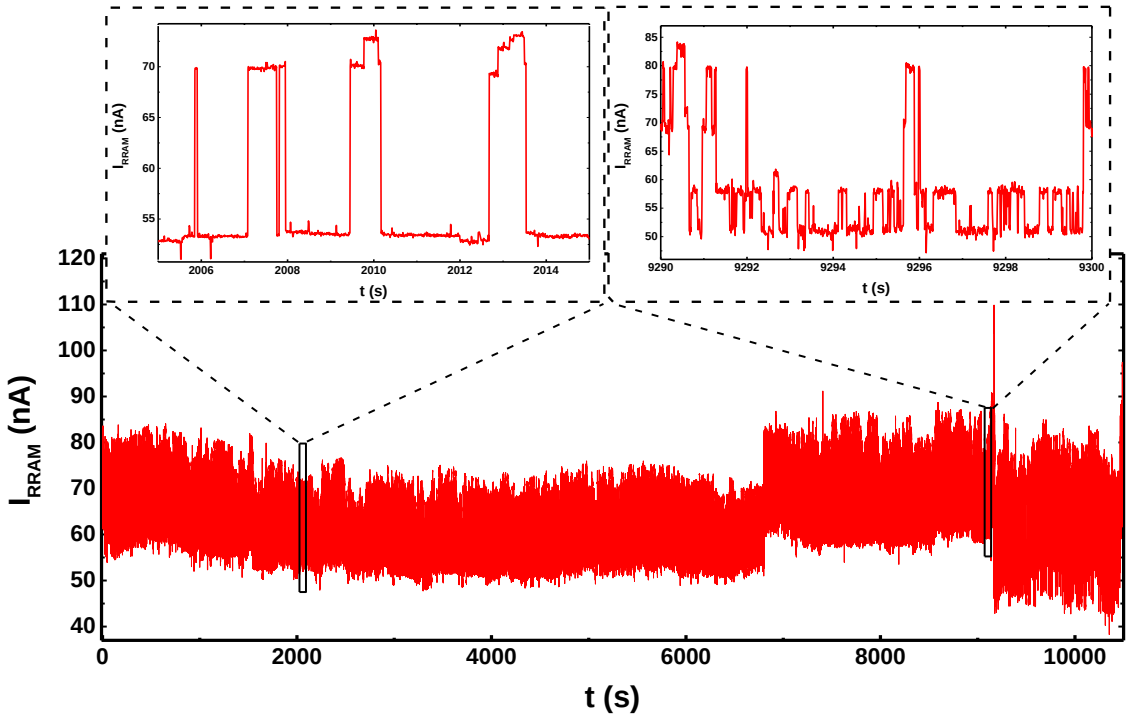


Figure 2: Current versus time for the HRS of Ni/HfO_2/Si devices at V_{RRAM}=-0.5V (more than 1.2 million data were measured in 10500 seconds). Zoomed in plots of RTN in two time windows are also shown.

3.- REVISION OF RTN ANALYSIS TECHNIQUES

The basic RTN signal representation consist of a current versus time plot. The trace consist of a series of points (I_i, t_i), where $i \in [1, N]$, being N the number of sampled data. In a basic plot it is easy to identify the number of current levels involved; in addition, a rough estimation of the more likely current states in a certain time window can be performed visually, see for instance Figure 3a (the description of the symbols employed in the Figure is presented in Table I) where a two-level RTN signal is shown in trace#640 (the original (I-t) trace was divided in 1203 intervals, from trace#1 to trace#1203). The current levels were labeled as states 0 and 1 (55nA and 72nA respectively), the 0 state being more stable than the 1 state. If the number of sampled data is high or the signal is

very complex, including many current levels, irreversible states, background noise, etc., this type of plots is not adequate since much of the statistical information is not described.

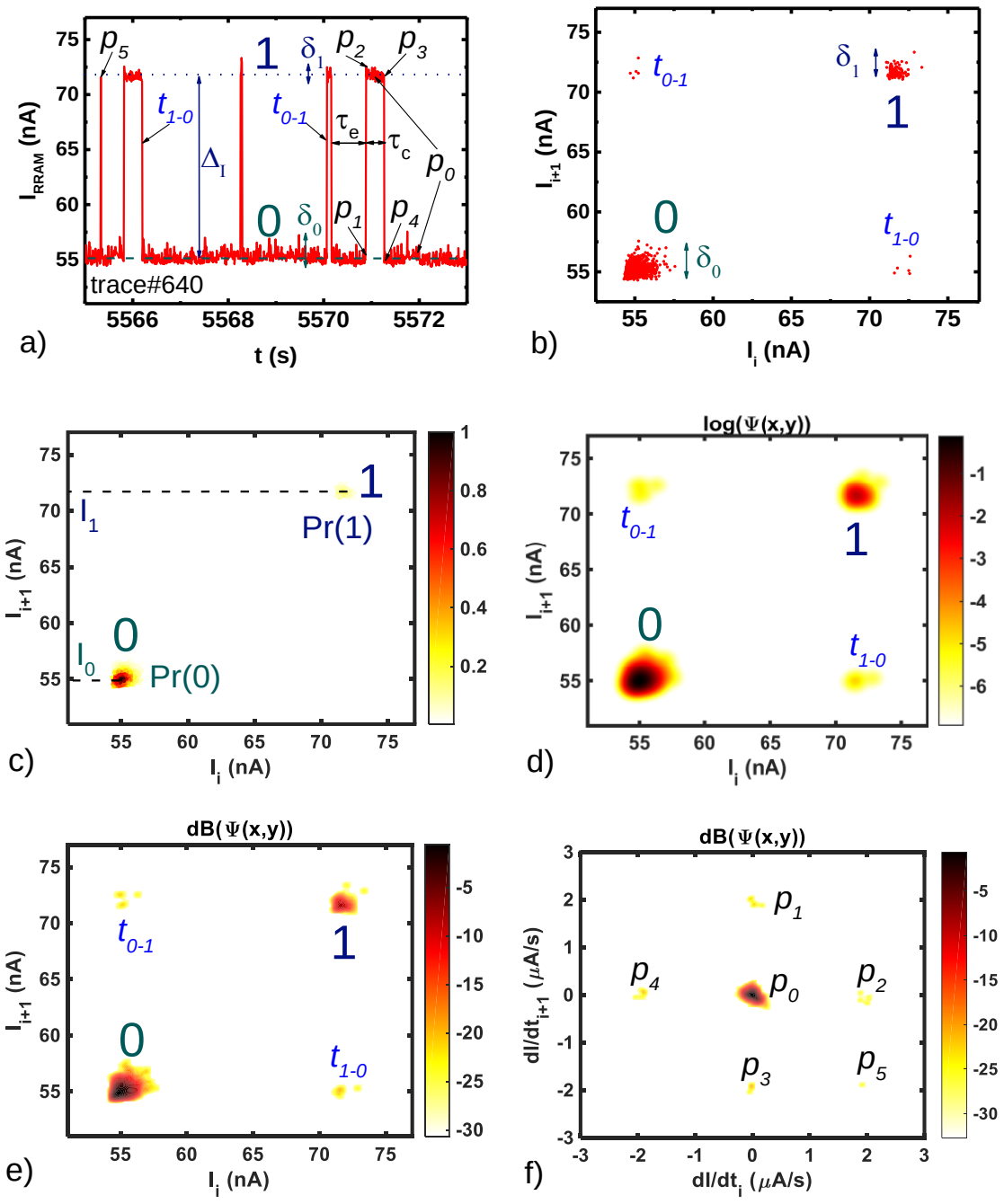


Figure 3. Different analysis methods for RTN signals. a) Current at V_{RRAM}=-0.5V versus time, trace #640 (1000 samples), b) Time lag plot (TLP) [31], c) Radius time lag plot (RTLP) [34], the most likely state is $Pr(0)$ presented by points in black, state 1 with a relative probability $Pr(1)$ close to 0.2 is shown in yellow,

d) WTLP (M=200 points, α=0.1) [32,33], e) LWTLP (M=200 points, M5x5, $\sigma_f = 1.5$) [25], f) dLWTLP (M=200 points, M5x5, $\sigma_f = 1.5$) [35].

A step ahead from the direct current trace plot is the Time Lag Plot (TLP), a representation method to easily visualize some RTN features. In this type of plot, the I_{i+1} current level versus the previous value, I_i, is plotted [31], see Figure 3b. The stable current levels appear as points clusters along the diagonal of the TPL graph ($I_{i+1} = I_i$), while transitions between the current levels show up outside this diagonal. Figure 3b shows the TLP of trace#640, where two clusters represent states 0 and 1 shown in Figure 3a. Few points outside the diagonal correspond to the transitions from state 0 to state 1, t_{0-1}, and the transitions from state 1 to 0, t_{1-0}. State 0 is described by a cluster with a greater number of points and also with greater dispersion than the cluster that describes state 1 ($\delta_0 > \delta_1$).

a.- An alternative to improve the TLP features, known as the Radius Time Lag Plot (RTLP) incorporates a method to visualize the likelihood of occurrence of a point in the TLP (Figure 3c). The value associated to a given position (I_i, I_{i+1}) depends on the number of counts in the neighborhood $\xi(i,j)$, within a region described by a given radius, r, [34]. In that manner, the areas with a higher point concentration are highlighted. The method can be described by Equations 1 and 2,

$$A\big(I(i), I(i+1)\big) = \sum_{j=1}^{N-1} \xi(i,j) \tag{1}$$

$$\xi(i,j) = \begin{cases} 1 & if \quad d\{(I(j), I(j+1)), (I(i), I(i+1))\} \leq r \\ 0 & if \quad d\{(I(j), I(j+1)), (I(i), I(i+1))\}) > r \end{cases} \tag{2}$$

where N is the number of samples (measured points in a (I-t) trace), d is the Euclidean distance function. The plot shows the same points as in the TLP (Figure 3b) but a colour map is employed to represent the number of points in the determined neighborhood (*Pr*, the relative probability normalized to unity is used).

Table I. Description of the symbols employed in Figure 3.

Mark	Description
1/0	stable states corresponding to the high/low current levels
$\boldsymbol{I_1}$ $(\boldsymbol{I_0})$	Mean current values corresponding to the high/low current levels
$\boldsymbol{\Delta_I}$	Difference between the high and low current levels
$\boldsymbol{t_{0-1}}$ $(\boldsymbol{t_{1-0}})$	Transition of low-high (high-low) current states
$\boldsymbol{\delta_1}$ $(\boldsymbol{\delta_0})$	Variability of the high (low) current states
$\boldsymbol{\tau_e}$ $(\boldsymbol{\tau_c})$	Emission (capture) times
$\boldsymbol{p_0}$	Current switch between stable current levels. It is obtained for stable levels, 0 or 1
$\boldsymbol{p_1}$	Beginning of the current transition from state 0 to state 1. The *i*-derivative is close to cero, while the following (*i+1*-derivative) is positive.
$\boldsymbol{p_2}$	End of transition revealed by p_1 (from 0 to 1). The *i*-derivative is positive, the *i+1* one is close to cero.
$\boldsymbol{p_3}$	Beginning of the current transition from state 1 to state 0. The *i*-derivative is close to cero, while the following (*i+1*-derivative) is negative.
$\boldsymbol{p_4}$	End of the transition revealed by p_3 (from 1 to 0 states). The *i*-derivative is negative, the *i+1*-derivative one is close to cero.
$\boldsymbol{p_5}$	Spike between states 0-1-0. It reveals a duration in the 1 state shorter than two sampling time steps.

b.- The Weighted Time Lag Plot (WTLP) calculates the occurrence probability in a given position of the TLP space considering the occurrences in the rest of positions in the graph by means of the bidimensional Gaussian distribution (BGD) [32, 33, 42]. It reduces the data space (I_i, I_{i+1}) of a trace to a $M \times M$ matrix ($M < N$); that is, for each position (x, y), the probability is calculated as a weighted sum of all the sampled data (I_i, I_{i+1}). The weights are given by the distance to each occurrence and the corresponding value of the BGD. The mathematical methodology can be described by Equation 3 [32],

$$\psi(x,y) = \frac{k}{2\pi\alpha^2} \sum_{i=1}^{N-1} \exp\left(\frac{-[(I_i - x)^2 + (I_{i+1} - y)^2]}{2\alpha^2}\right) \quad (3)$$

The values of x and y depend on the range of values to represent and the numbers of points in the matrix accordingly to Equation 4.

$$(x, y) = \left(\frac{\hat{x}}{M-1}(I_{max} - I_{min}) + I_{min}, \frac{\hat{y}}{M-1}(I_{max} - I_{min}) + I_{min}\right) \quad (4)$$

where $\hat{x}$ e $\hat{y}$ are the indices of the matrix of values to be displayed ($\hat{x} = 0, 1, \ldots M-1$; $\hat{y} = 0, 1, \ldots M-1$), and I_{max} e I_{min} are the maximum and minimum current values used in the plot. Figure 3d shows the WTLP of the trace #640 in logarithmic scale with a value of M=200. The state with the higher probability of occurrence can be easily visualized in the main diagonal making use of a color map. Both methods (RTLP and WTLP) are computationally expensive.

c.- The LWTLP reduces the space (I_i, I_{i+1}) to $M \times M$ points (x, y) [25], like the method WTLP (Equation 4 to discretize the space (I_i, I_{i+1}) can be also used here). For the WTLP case, instead of calculating for each point (x, y) the contribution of all the points (I_i, I_{i+1}) weighted by BGD, the coordinates $(\hat{x}, \hat{y})$ are determined for each of the $N-1$ pairs of points (I_i, I_{i+1}) by the following equation:

$$(\hat{x}, \hat{y}) = \left(int\left(M\frac{I_i - I_{min}}{I_{max} - I_{min}}\right), int\left(M\frac{I_{i+1} - I_{min}}{I_{max} - I_{min}}\right)\right) \quad (5)$$

where $int(x)$ is the function that returns the integer closest to x. For the occurrence probability calculation, different submatrices were used (for example 3×3 or 5×5) [25]; a 5x5 matrix example is shown in Figure 4 with the corresponding weights.

0.16	0.33	0.41	0.33	0.16	$\hat{y}$+2
0.33	0.64	0.80	0.64	0.33	$\hat{y}+1$
0.41	0.80	1	0.80	0.41	$\hat{y}$
0.33	0.64	0.80	0.64	0.33	$\hat{y}$-1
0.16	0.33	0.41	0.33	0.16	$\hat{y}-2$
$\hat{x}$-2	$\hat{x}$-1	$\hat{x}$	$\hat{x}$+1	$\hat{x}$+2	

Figure 4: An example of LWTLP 5x5 matrix is shown, equal weights are plotted with the same color [25].

Figure 3e shows the LWTLP of trace #640 calculated for M=200 points and a 5x5 sub-matrix. The main virtue of this method is its computational efficiency, approximately 750 times faster than WTLP or RTLP [25].

d.- The Differential Locally Weighted Time Lag Plot (DLWTLP) is a new method of RTN representation [35] based on the numerical derivative (Equation 8) of the current with respect to time, instead of the current itself. This plot does not represent directly the RTN signal stable levels but offers information about the temporal variation of current values, such as the probability of occurrence of changes between current states (p_1, p_2, p_3 and p_4) and the presence of spikes (p_5), as described in the Table I [35].

$$\frac{dI_i}{dt} \approx \frac{I_{i+1} - I_i}{t_{i+1} - t_i} \ , i = 1..N - 1. \tag{8}$$

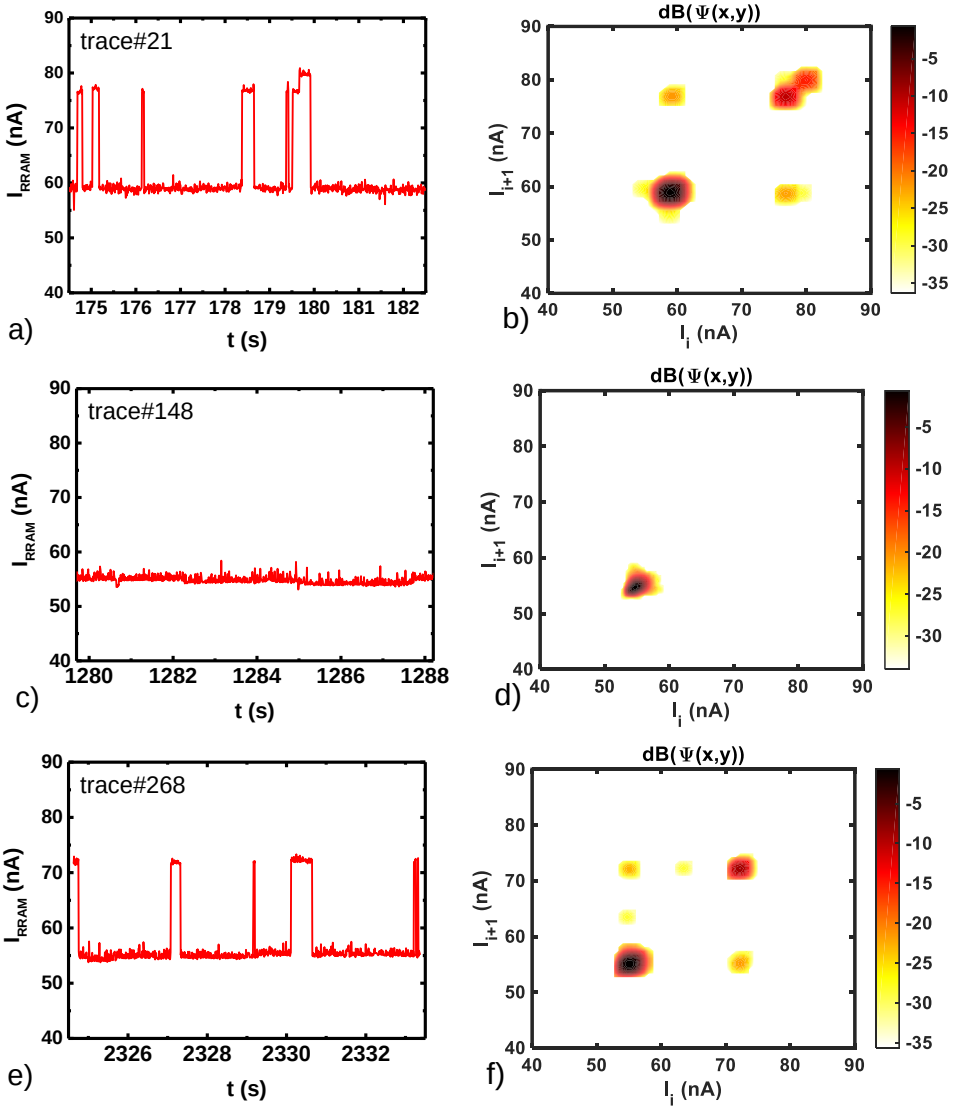
Figure 3f shows the DLWTLP of trace #640, the occurrence of p_5 indicates the presence of spikes in the RTN signal. The marks p_1 and p_2 indicate the transition from state 0 to state 1, and the presence of marks p_3 and p_4 indicates the transition from state 1 to state 0.

4.- NEW TECHNIQUE BASED ON NEURAL NETWORK ANALYSIS

4.1.-Study of characteristic patterns in the LWTLP domain for different RTN signals

Among the methods described in the previous section, the LWTLP allows to characterize the RTN signal in a determined (I-t) trace with a low computational cost. This computation efficiency becomes evident when we need to analyze long RTN traces in comparison with other methodologies. Figure 5 shows the LWTLP (M5x5 and M=50 points) for different types of RTN signals found in the long trace in Figure 2. Figure 5a shows trace #21 (multilevel fluctuation with three current levels) and Figure 5b the corresponding LWLTP; Figure 5c shows trace #148 (representing background noise), Figure 5d shows the related

LWTLP. Figure 5e represents trace #268 (a stable two current level fluctuation), Figure 5f plots the LWTLP for trace #268. Figure 5g shows trace #1074 (a multilevel (6 levels) non correlated transition fluctuation) and Figure 5h the corresponding LWTLP. It is revealed that different types of RTN signals show characteristic representations in terms of LWTLP patterns. The LWTLP patterns can be seen as a representation of RTN in a different domain that highlights important signal features with respect to the temporal domain, just as Fourier domain allows for the frequency analysis of a certain temporal signal.



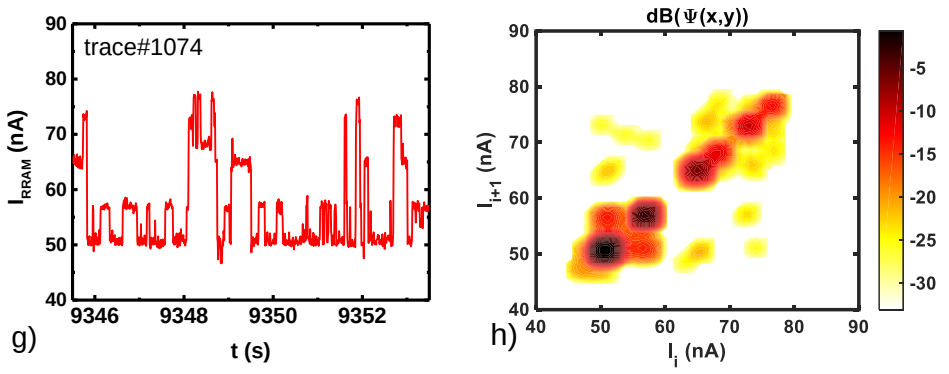


Figure 5: Examples of different types of RTN signals found in the long trace represented in Figure 2. Comparison of (I-t) trace time intervals and the corresponding LWTLPs (with M5x5 and M=50 points [25]). a) Trace #21 (multilevel fluctuation with 3 current levels), b) LWTLP of trace #21; c) Trace #148, (background noise), d) LWTLP of trace #148; e) Trace #268 (stable two current level fluctuation), f) LWTLP of trace #268; g) Trace #1074 (multilevel (6 levels) non correlated transitions fluctuation), h) LWTLP of trace #1074.

The long-duration signal in Figure 2 (around 1.2 million sampled data) is divided into 1203 pieces (m=1203), for each of these pieces the LWTLP (M5x5 M= 50 points) is obtained. These LWTLP plots are the input set to train an artificial self-associative neural network. In SOM, each neuron plays the role of a cluster head and is connected to all the input variables. During training, neurons adapt the weights that connect them to inputs. Specifically, a similarity measure-based algorithm is used to determine the most similar neuron or cluster head. This neuron and their neighboring nodes slightly modify their own weights in order to increase the similarity with that particular input in the training process. The strength of the adaptation process decreases both with distance (the effect is higher on closer neurons) and time (the last training examples have less influence on the map than the first examples). After training, the patterns are clustered by mapping them to the most similar neuron. One advantage of using SOM with respect to well-known methods, such as k-means [39], and which motivated our choice, is the preservation of the topology. In our application, it is interesting to know that patterns corresponding to a certain neuron will be similar to patterns corresponding to the neighboring nodes and visualize this information in the 2D map. See the Appendix for further details.

Several network sizes have been tested for this purpose and we found that a 5x4 neurons network is an acceptable compromise between the number of clusters or classes obtained and the elements classified in each of the classes. In the training process 1203 input vectors

(2500 dimension) were used for their training employing 4000 epochs (iterations for learning performed by the network, updating the weight W_i), using 31 minutes in this process (for an Intel ® Core™ i7-7700HQ CPU @2.8GHz and 24GB RAM). The graphic representation of the neural network is presented in Figure 6. The degree of connection of each neuron with its neighbors is represented by a color scale (dark black furthest, clear-yellow closer).

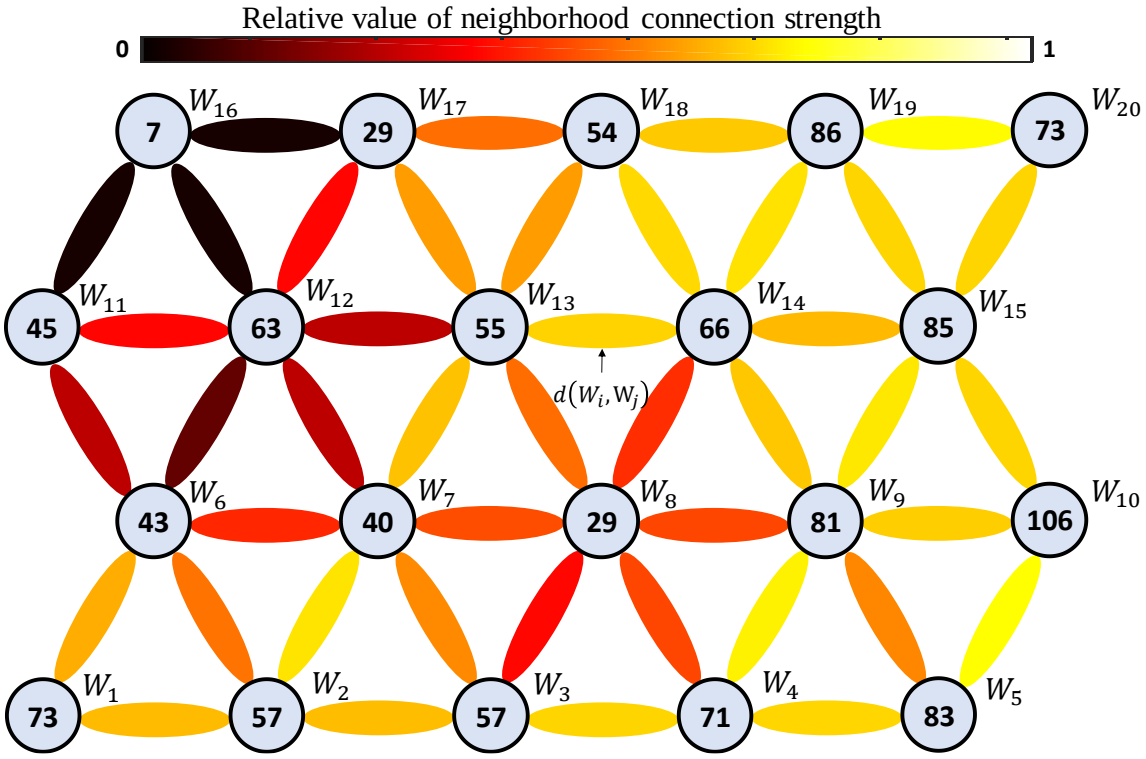


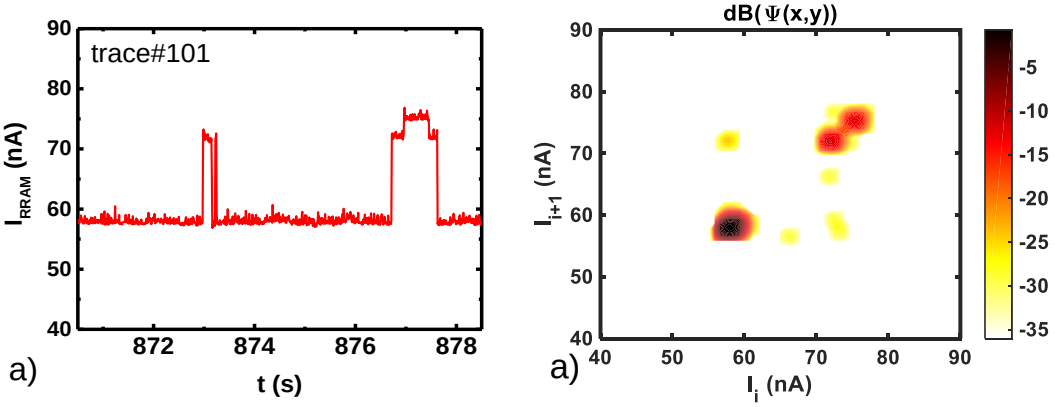
Figure 6. Representation of the SOM 5x4 neural network employed for our analysis. The circles represent the neurons, the number within the circles stand for the number of LWTLP patterns associated to each clusterhead (neurons). The neighborhood connection strength of the neurons in the self-associative network is represented in colour (a dark colour connection between the neurons corresponds to a large Euclidean distance between them, i.e., Euclidean distance of the corresponding neurons weights $d(W_i,W_j)$. On the other hand, a light colour between the neurons connection means that neurons are close to each other in terms of Euclidean distance). A two-dimensional mapping is employed for the implemented ANN, we have chosen this particular topology although others can be employed.

The neural network provides a total of 20 clusters (C#1 to C#20) to classify each of the 1203 input vectors (RTN traces), notice that the sum of the number of patterns associated to the 20 different clusters represents the 1203 traces. Figure 6 represents the number of

input vectors that have been classified in each cluster, we have clusters with many associated vectors (C#10 =106) and others sparsely populated (C#16 =7).

5.- RESULTS AND DISCUSSION

The main result of the new procedure introduced here can be found at the output of the ANN training process. A set of clusters that allow the classification of the RTN signals experimentally measured is obtained. The pattern characteristics associated to each cluster represent a particular configuration of the RTN signal in terms of number of current levels, their corresponding likelihood of occurrence, current levels transitions, background noise, etc. In order to shed light on this issue, we present three examples of the patterns associated to the cluster C#1, C#11 and C#16, obtained after the ANN training process. They are shown in Figure 7, C#1: a) trace #188, b) trace #101, c) trace #868; Figure 8, C#11: a) trace #1061, b) trace #1069, c) trace #1086; and Figure 9 C#16: a) trace # 148, b) trace # 427, c) trace # 553). In these figures we can observed the (I-t) traces and the corresponding LWTLPs (calculated with M5x5 and M=50 points).



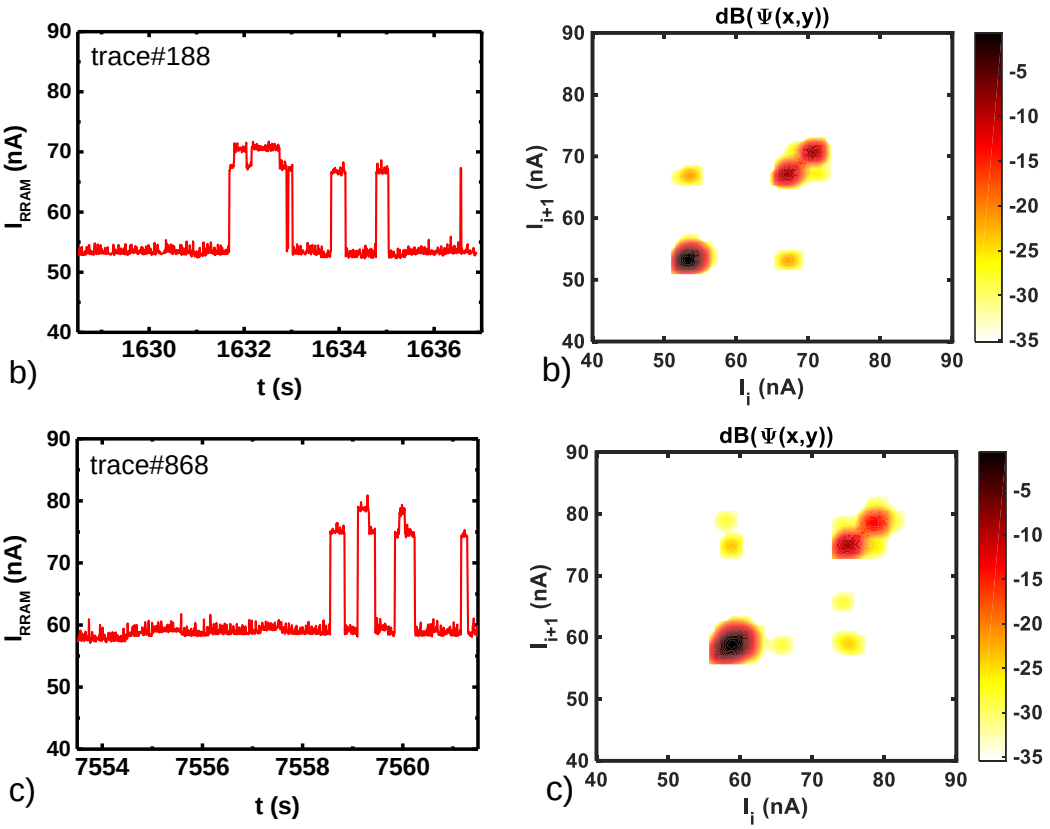


Figure 7. I-t trace plots and corresponding LWTLPs (calculated with M5x5 and M=50 points) classified in the same cluster C#1: a) trace #188, b) trace # 101, c) trace # 868.

In Figure 7 we deal with signals that show three current levels. Notice that the traces considered are just different parts of a long RTN trace (Figure 2), these traces corresponding to different time windows. The three corresponding dark clusters of points in the LWTLPs plots are clearly detected. In all cases we see this behavior, although we can distinguish certain differences in the probability of occurrence of each of these current levels.

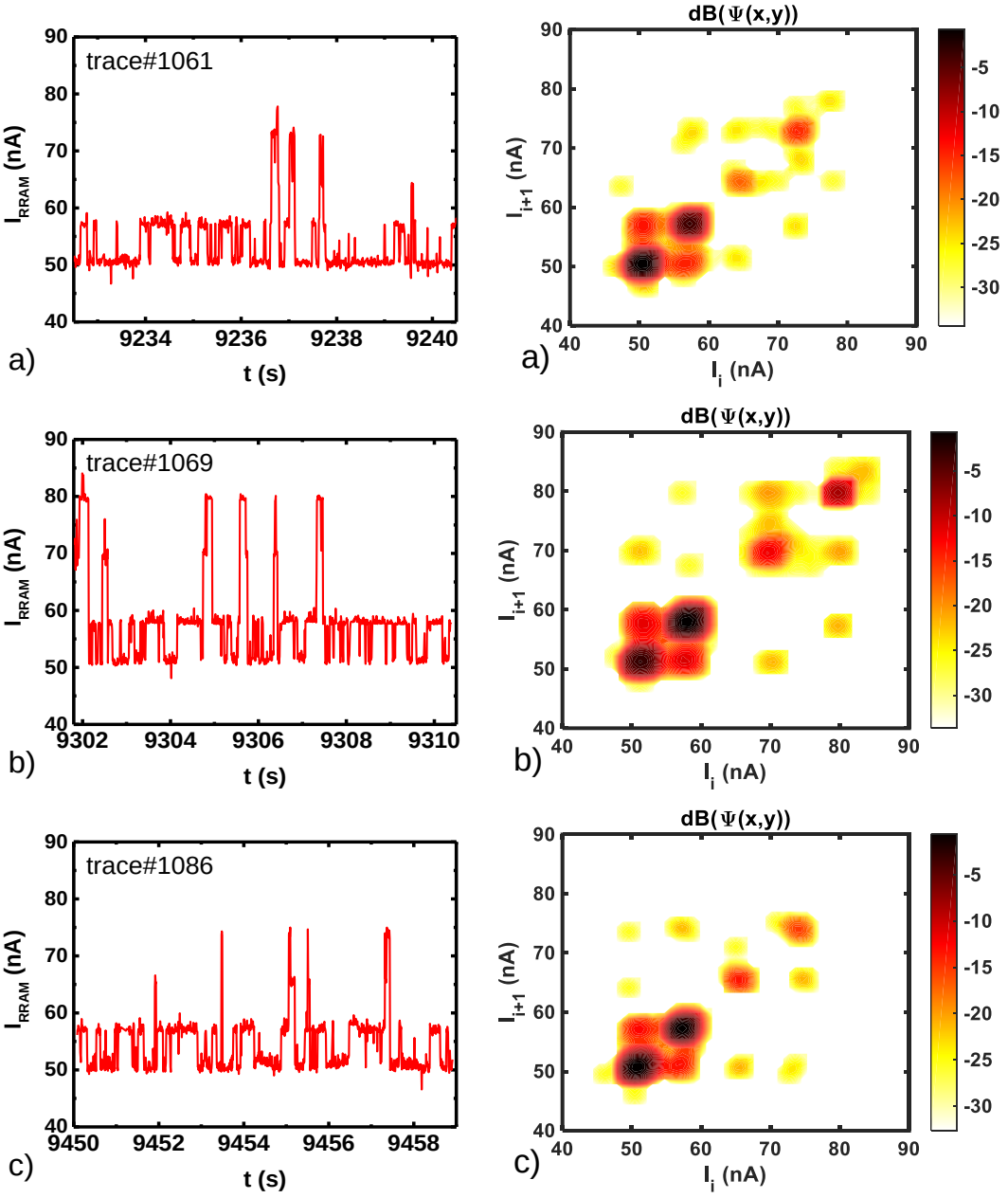


Figure 8. I-t trace plots and corresponding LWTLPs (calculated with M5x5 and M=50 points) classified in the same cluster C#11: a) trace #1061, b) trace # 1069, c) trace # 1086.

In Figure 8 the signals associated to cluster C#11 show four current levels, the LWTLP patterns clearly indicate this fact, although variance in the probability of occurrence is observed. The traps linked to the current levels seem to be independent and additive since the patterns in the LWTLP domain are stable and independent.

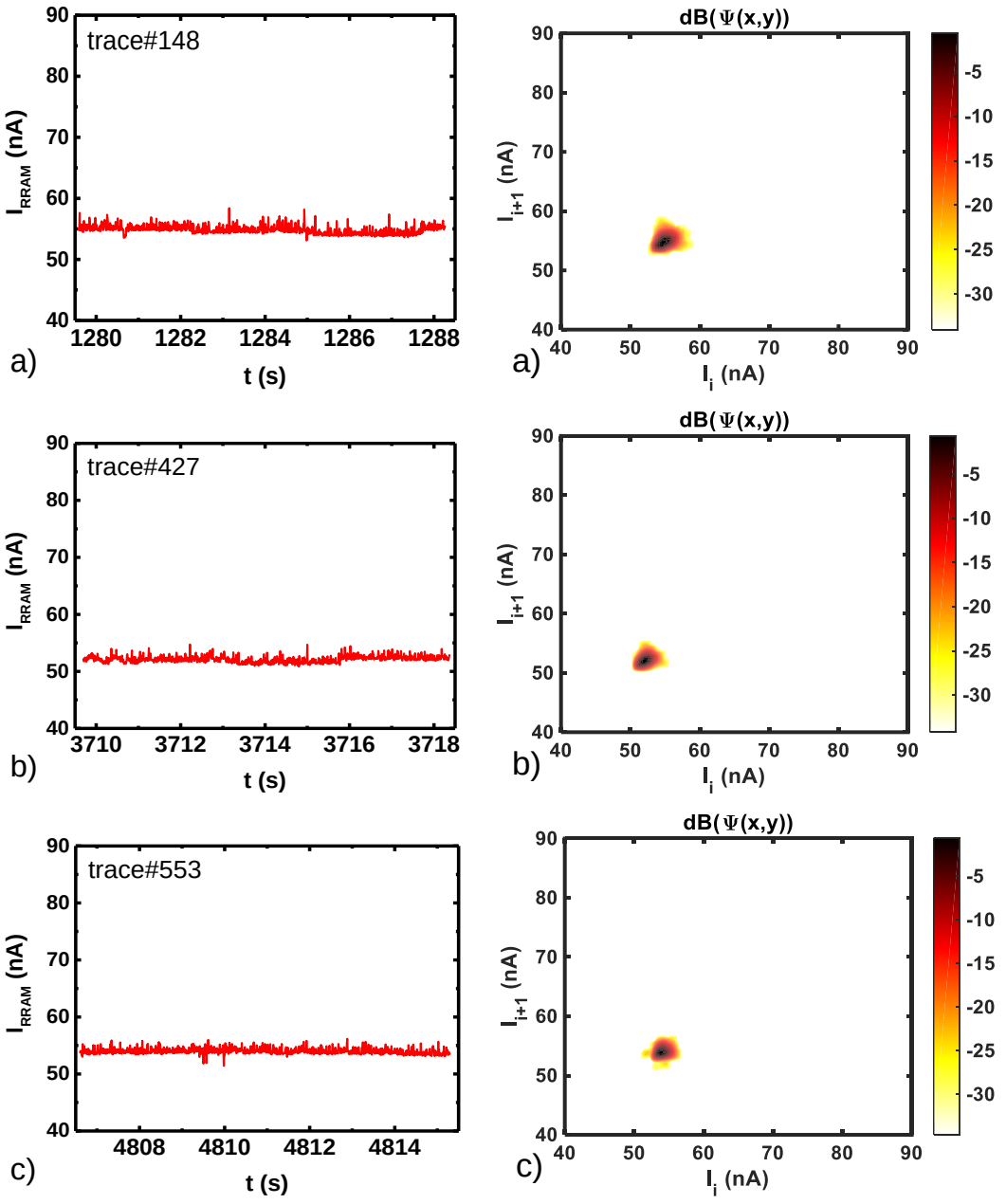


Figure 9. I-t trace plots and corresponding LWTLPs (calculated with M5x5 and M=50 points) classified in the same cluster C#16: a) trace #148, b) trace #427, c) trace # 553.

In Figure 9 the representations indicate that we are facing background noise. No clear current levels can be appreciated. As expected, a single cluster of points at the LWTLP domain center is obtained in all cases.

According to the obtained results, for the neuron set selected, the most representative LWTLP patterns (and consequently the RTN signals) are classified after the training

process. In order to highlight this point, one LWTLP pattern (footprint) for each of the 20 clusters produced (C#1 to C#20) is shown in Figure 10. A general view of this figure allows to summarize the main features of the RTN signal plotted in Figure 2.

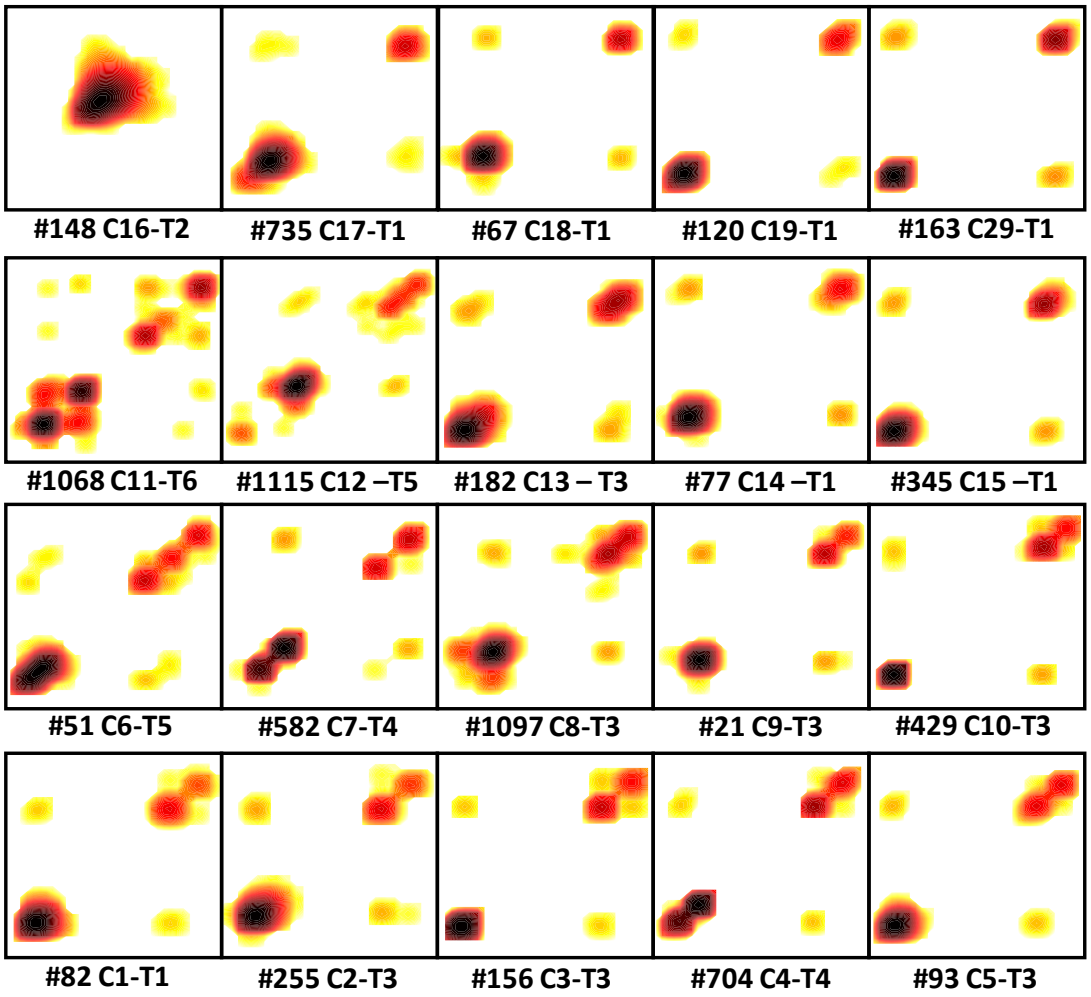


Figure 10: LWTLP pattern types for each of the 20 clusters obtained (C1 to C20) after SOM training process. The classification of the obtained RTN signal reference patterns is described in Table II.

These results are comprehensively classified in Table II, the classification of the LWTLP patterns is done by considering six different RTN signal reference types (T1 to T6). The corresponding probability for these reference types is also given for comparison purposes.

Table II. LWTLP pattern classification based on RTN fluctuation types.

Type of RTN fluctuation			Clusters #	RTN Type	%
Two-Levels	Stable-Current		14,15,17,18,19,20	T1	32,67
	Background noise		16	T2	0,58
Multilevel	Non correlated transitions	3-levels	1,2,3,5,8,9,10,13	T3	44,97
		4-levels	4,7	T4	9,23
		5-levels	6,12	T5	8,81
	Interactive defects		11	T6	3,74

Finally, the evolution of the reference fluctuation types (shown in Table II) along the RTN trace under consideration is represented in Figure 11. It can be observed that the most probable LWTLP pattern corresponds to the reference type T3 (44.97%) followed by reference type T1 (32.67%). Reference type T3 corresponds to a RTN signal with three current levels. This pattern may be explained by the activation of two defects, one generates two different current levels and the other produces other two levels, one of them merges with one of the levels of the companion defect. Reference type T1 corresponds to a two level stable current fluctuation produced by a single active defect. These are the two most probable situations. In relation to reference type T6, it is important to highlight that interactive defects refer to defects whose fluctuation may involve changes in the occupation probability of other traps, resulting in a large distribution of switching times and RTN amplitudes [36]. In our calculations, T6 is unlikely although it can occur. T4 and T5 can be attributed to two or three active defects.

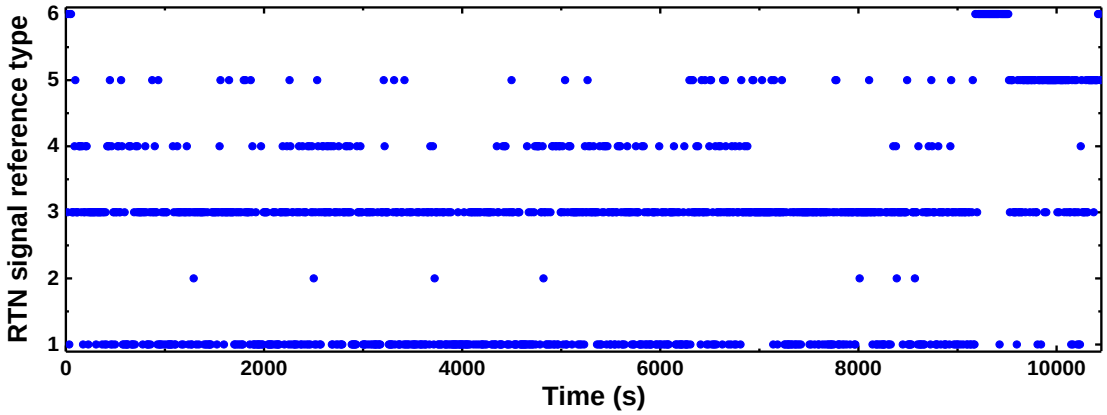


Figure 11. Evolution of the reference fluctuations types (T1-T6) employed to classify the RTN signals corresponding to the LWTLP patterns shown in Figure 10.

An interesting issue is related to the ability of the method to track the probability that a transition type is reached from another type. The evolution of the RTN signal reference types (T1-T6) shown in Figure 11 allows the determination of the state transition matrix that represents the count of observed transitions from state i to state j (i and j ϵ [T1, T6]) and can be represented graphically as a Discrete-Time Markov Chain, see Figure 12. This latter mathematical tool has been previously employed in the analysis of RTN signals [43].

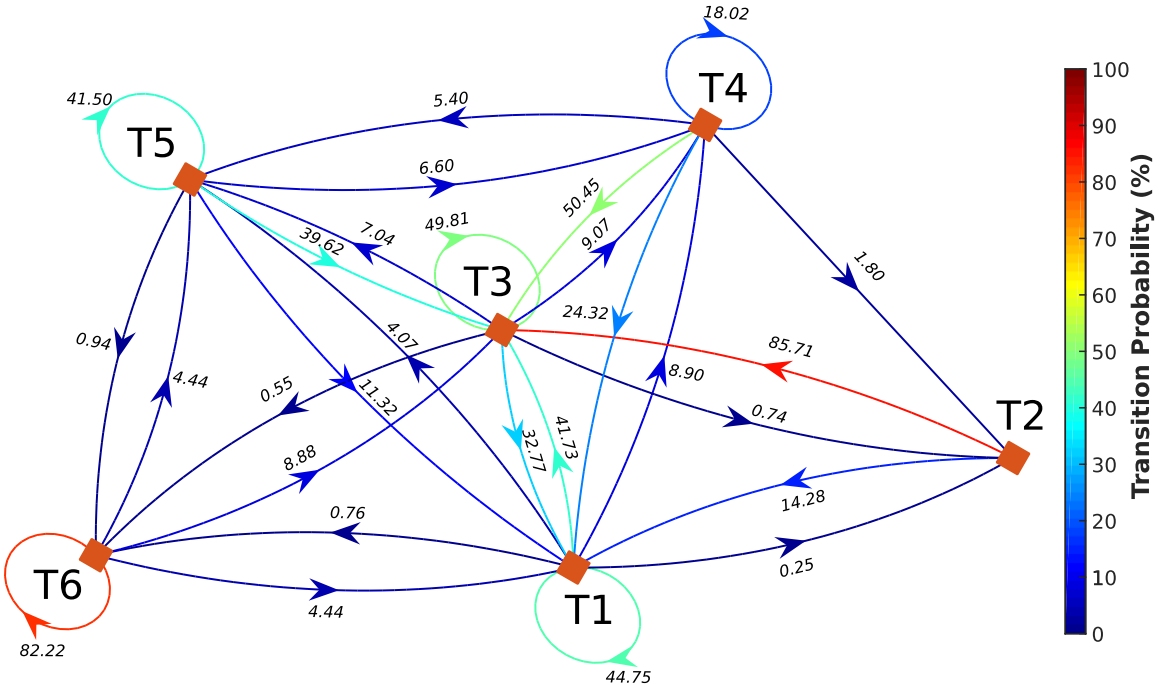


Figure 12: Discrete-Time Markov Chains extracted from evolution of the fluctuations reference types of Figure 11. Notice that we have used a non-normalized transition matrix, the transition probabilities are multiplied by 100 to ease the interpretation with percentages.

Several conclusions can be drawn from the interpretation of Discrete-Time Markov Chains. Being in state T1, it is most likely to remain in T1 (44.75%) and the most probable transition from T1 is to T3 with a probability of 41.73% (this would correspond to the activation of an additional defect in addition to one currently active). Most likely, T1 will be preceded by T3 (32.77%) or T4 (24.32%) and these two latter transitions could correspond to the deactivation of a defect. The transition from T2 (14.28%) may be due to

the activation of a defect in the case none were active and the transition from T5 (11.32%), corresponds to the deactivation of two defects, which has less likelihood as expected. The probability of reaching T2 is low and the most probable case would be preceded by T4 (1.8%). From T2 we will most likely end up in T3 (85.71%) as well as it is not likely to remain in T2. Other transitions can be easily described by using Figure 12. The knowledge of the state transitions and the occupancy probability of RTN reference types can be employed to deepen physically speaking on the nature of RTN signals. This is linked to the number and type of active defects in each time window. In addition, this method correlates the nature of electrically active defects with the trap response at the device level. This correlation allows to define optimal design rules to keep a defect control and optimize the device performance. The Markov chain transition matrix can also help on facets such as the use of RRAMs as entropy sources for random number generators and physical unclonable functions, used in applications such as stochastic computing or hardware security [44].

It is also interesting to comment on the issue linked to the time intervals employed to perform this study since these time intervals can be chosen at will for a determined long experimental (I-t) set of data. The length of the time intervals could affect the results obtained by the methodology presented here that is why we analyze it. The number of traces (denoted by m in previous sections) would change depending on the number of sampling points (NSP) for the time windows selected (N, as we denoted it Section 2). This choice would influence the LWTLPs obtained; therefore, there could be changes in the types of clusters obtained (Figure 10) and in the time linked to the RTN signal reference types (last column of Table II). We have tried to perform a study to clarify the impact of the NSP on the obtained results. In Fig. 13, experimental current values for a determined time interval corresponding to N=1500 (Figure 13a) are shown. It can be observed that depending on N in this particular case (i.e., the length of the time window), the associated LWTLPs are different since certain current levels can be included (or not) and this is reflected upon the LWTLP.

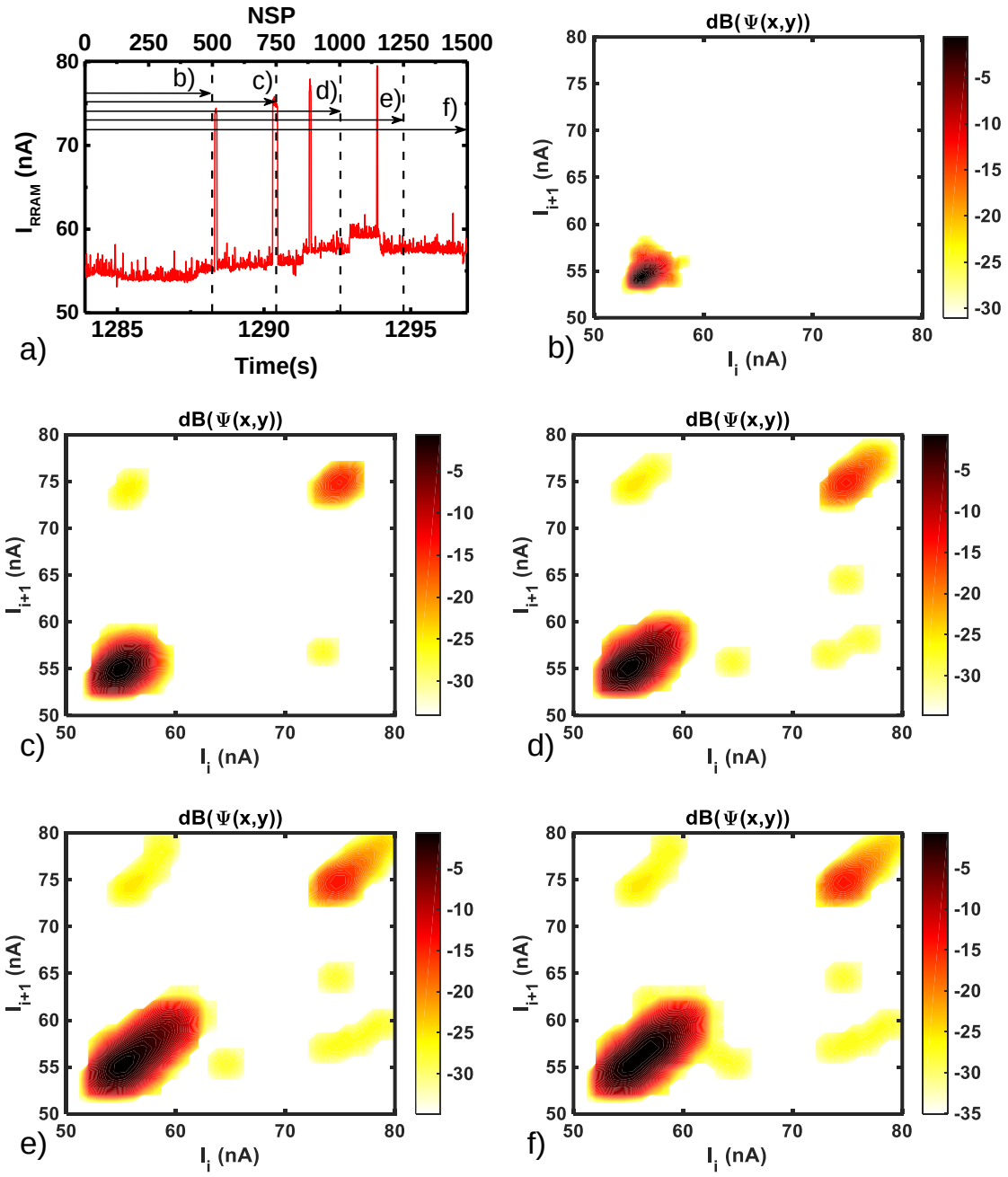


Figure 13. (a) Current versus time for the studied time intervals of measured data indicated by (b), (c), (d), (e) and (f). LWTLPs for time intervals with different N, b) 500 sampling points, c) 750 sampling points, d) 1000 sampling points, e) 1250 sampling points and f) 1500 sampling points.

The whole study for intervals with different N has been performed, where no significant differences were observed regarding the clusters shown in Figure 10. With respect to the

time associated to the different RTN fluctuation types, the results have been presented in Table III.

Table III. Time expressed in percentage associated to the RTN fluctuation types obtained in the SOM training process for intervals with different number of sampling points.

	N = 500	N = 750	N = 1000	N = 1250
T1 (%)	44.20	39.72	32.67	30.61
T2 (%)	6.48	1.01	0.58	-
T3 (%)	41.71	43.82	44.97	47.67
T4 (%)	4.53	7.82	9.23	8.69
T5 (%)	1.45	4.29	8.81	9.31
T6 (%)	1.62	3.34	3.74	3.72

Although some differences can be appreciated for the shorter time intervals, there is a saturation effect for N > 750. In this case no significant changes are obtained in the results, as can be seen in Table III. We selected N=1000 both to have a higher number of traces (m) than for N>1000, and consequently increase the number of input vectors for the ANN training process, and because of the results obtained in Table III show no significant differences to the case N=1250. In these considerations it is the optimum N value.

5.- CONCLUSIONS

A new technique based on ANN has been implemented to analyze RTN signals. The procedure takes advantage of the analysis of SOMs based on the data obtained by means of the LWTLP technique for sets of experimental RTN data taken from a long (I-t) trace. In particular, we can graphically characterize by the LWTLP sequence of RTN data (for instance, 1000 sampled current versus time data) with a high processing speed by means of a pattern (50x50 points) set. With these pattern set, making use of its vector form, we train a SOM neural network (in our case a network made of 5x4 neurons) that allows us to classify each of the 1203 input sequences (RTN signals for a determined time window) in each of the 20 patterns that have been generated by the neural network training process. An individualized analysis of each of the 20 patterns allows the classification of the

different types of current fluctuations present in a long sequence of RTN data, more than 1.2 million sampled data for our particular measurement. Once these results have been achieved, the graphical representation of the type of fluctuations that can be found at each time window in the complete long RTN signal can be determined. Finally, we concluded that six types of patterns were present in the RTN signal trace we analyzed, including stable 2-level current fluctuations, multilevel transitions and the presence of background noise.

6.- ACKNOWLEDGMENTS

The authors thank the support of the Spanish Ministry of Science and Universities and the FEDER program through projects TEC2017-84321-C4-1-R, TEC2017-84321-C4-3-R, TEC2017-84321-C4-4-R. This work has made use of the Spanish ICTS Network MICRONANOFABS.

7.- APPENDIX

Self-organizing maps (SOM), also namely as Kohonen networks [37, 38], are accounted among the artificial neural networks (ANN). They are bio-inspired software models based on neuron organization into the human brain [Ch. 4, 45]. In this work, SOMs are used for the task of clustering patterns as a visual alternative to other well-known approaches such as k-means [39]. Note that in the context of deep learning [46], other unsupervised approaches such as the Restricted Boltzmann Machine (RBM) [47], Variational Autoencoders (VAEs) [48] or generative adversarial networks [49] among others go beyond the clustering task and try to learn a probabilistic model of the dataset.

The training process linked to SOMs makes use of an unsupervised learning algorithm, it allows to analyze and visualize high-dimensional data into a low-dimensional discretized space (they are mostly employed in a two-dimensional domain, the designated map) that represents the input space of the training samples. Two main SOM features make them different from other widely used ANNs in supervised or semi-supervised tasks, such as the multi-layer perceptron, convolutional neural networks or recurrent neural networks [46]. On the one hand, they are competitive neural networks that implement the winner-take-all function; on the other hand, they have a neural network plasticity that modifies the local synaptic weight as a function of the neighborhood related data.

The training expression for a neuron i with weight vector $W_i(e)$ that we have employed here is the following:

$$W_i(e+1) = W_i(e) + \Omega(i,j,e) \cdot \alpha(e) \cdot [LWTLP(t) - W_i(e)] \begin{cases} \forall i = 1 \dots n \\ j \in \{1..n\} \\ \forall t = 1 \dots m \end{cases} \quad \text{(A.1)}$$

$$\Omega(i,j,e) = \begin{cases} 0 & if \quad d_{i,j} > d \\ 1 & if \quad d_{i,j} \leq d \end{cases} \quad \text{(A.2)}$$

where e is the step training index, i is the index of the actual neuron weight vector to update $W_i(e+1)$, $LWTLP(t)$ is the vector input characteristic of each RTN signal trace, j is index of the best matching unit (BMU) that represents the neuron whose weight vector $W_j(e)$ is most similar to the input vector $LWTLP(t)$, t is the index of the input vector

$LWTLP(t)$, m is the number of input vectors $LWTLP(t)$. $\alpha(e)$ is a monotonically decreasing learning coefficient (for example, we could use this analytical description $\alpha(e) = \alpha_0/(1 + k \cdot e)$, where, α_0 is the initial learning coefficient and k is a parameter, $k > 0$), $\Omega(i, j, e)$, as a function of variable e, represents the neighborhood function which gives the distance ($d_{i,j}$) between the neuron i (actually updated neuron $W_i(e + 1)$) and the best matching unit j. If the neuron i is near neuron j ($d_{i,j} \leq d$) then the neuron i is updated (Equation A.2). Variable n is the number of neurons in the network (usually $n = x \cdot y$, where x and y are integers; in this respect, although we have employed 10·10, 6·6, 5·5 topologies, we found that an optimum choice was 5x4, see Figure A1).

The training utilizes a competitive learning method to update the weight of the network $W_i(e + 1)$. When an input training vector $LWTLP(t)$ is fed to the neurons network, W_i ($\forall i = 1 \ldots n$), its Euclidean distance to all weight vectors is computed. Neuron j, whose weight vector, $W_j(e)$, is the most similar to the input vector $LWTLP(t)$ (BMU), is determined. The BMU weights and neurons close to it in the SOM grid are recalculated taking into consideration the input vector (in our particular case, as explained in section 4, we have m=1203 input vectors and 4000 epochs). The magnitude of the changes in $W_i(e + 1)$ decreases with each training step (depending on function $\alpha(e)$) and with the grid-distance too (function of $\Omega(i, j, e)$) [50].

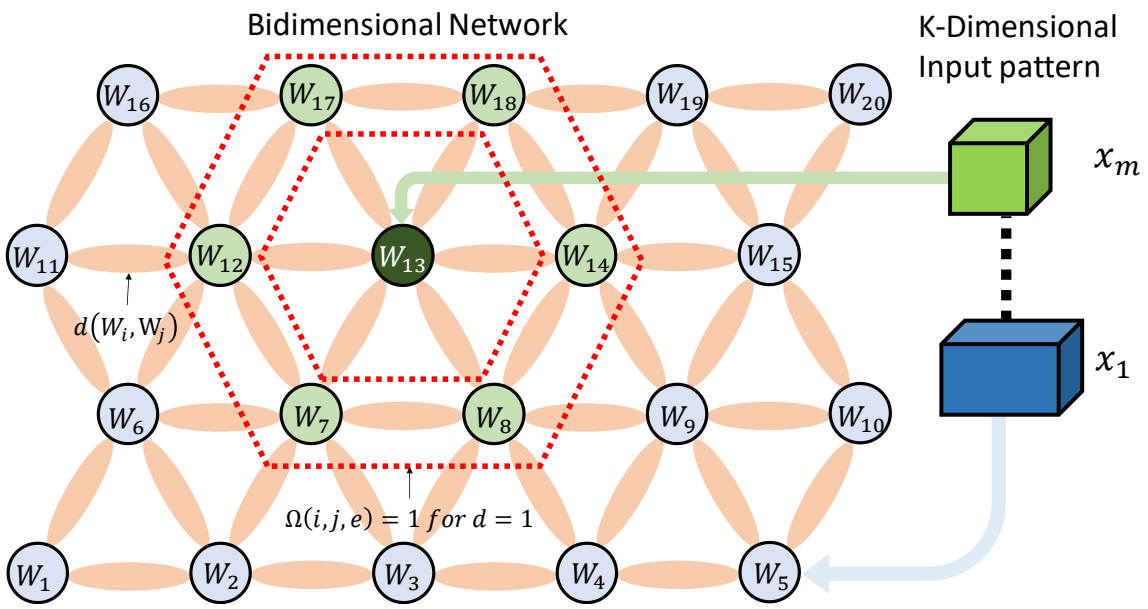


Figure A1: Example of SOM neighbors connections on a 5x4 neurons network. The connection of the neurons is done in the form of a bee panel (hexagonal arrangement), the number of inputs is an array of m values of K dimension (in our application m=1203 and K=2500). Each input pattern is assigned to the most similar neuron W_i. In training time the weight of each neuron is updated by equation A.1 (taking into account the neighborhood function described in Equation A.2). When the training process is finished, each new input pattern produces the activation of only one neuron of the network (for example, input pattern x_m produced the excitation of neuron W_{13}, the output signal of this is 1) while the rest of the neurons in the network are deactivated (producing a null output signal), the index of the active neuron determines the cluster or class assigned to the input.

The input $x(t)$ of the SOM is one K dimensional vector, the matrix LWTLP $(M \times M)$ is transformed in a K dimensional vector ($K = M \times M$), concatenating each row LWTLP matrix to the following one.

These are the ANN we have employed here to perform the analysis presented in this work. The results consisted of a set of clusters (where each neuron is a cluster head) obtained by using SOM with the goal to classify LWTLP patterns and characterize RTN signals.

References

[1] F. Pan, S. Gao, C. Chen, C. Song, F. Zeng, "Recent progress in resistive random access memories: materials, switching mechanisms and performance", Materials Science and Engineering, 83, pp. 1-59, 2014.

[2] M. Lanza, H.-S. P. Wong, E. Pop, D. Ielmini, D. Strukov, B.C. Regan, L. Larcher, M.A. Villena, J.J. Yang, L. Goux, A. Belmonte, Y. Yang, F. M. Puglisi, J. Kang, B. Magyari-Köpe, E. Yalon, A. Kenyon, M. Buckwell, A. Mehonic, A. Shluger, H. Li, T.-H. Hou, B. Hudec, D. Akinwande, R. Ge, S. Ambrogio, J.B. Roldan, E. Miranda, J. Suñe, K.L. Pey, X. Wu, N. Raghavan, E. Wu, W.D. Lu, G. Navarro, W. Zhang, H. Wu, R. Li, A. Holleitner, U. Wurstbauer, M. Lemme, M. Liu, S. Long, Q. Liu, H. Lv, A. Padovani, P. Pavan, I. Valov, X. Jing, T. Han, K. Zhu, S. Chen, F. Hui, Y. Shi, "Recommended methods to study resistive switching devices", Advanced Electronics Materials, 5, 1800143, 2019.

[3] S. Menzel, P. Kaupmann, R. Waser, "Understanding filamentary growth in electrochemical metallization memory cells using kinetic Monte Carlo simulations", Nanoscale, 7, 12673, 2015.

[4] M.A. Villena, J.B. Roldán, F. Jiménez-Molinos, E. Miranda, J. Suñé, M. Lanza," SIM2RRAM: a physical model for RRAM devices simulation", Journal of Computational Electronics, 16, p. 1095, 2017. https://doi.org/10.1007/s10825-017-1074-8

[5] D. Ielmini, R. Waser. "Resistive Switching: From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications", Wiley-VCH, 2015.

[6] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, H.-S. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation", IEEE Trans. Electron Devices, 58 (8), pp. 2729–2737, 2011. doi: 10.1109/TED.2011.2147791.

[7] S. Yu, “Neuro-inspired computing using resistive synaptic devices”, Springer, 2017. ISBN: 978-3-319-54312-3

[8] J. Zahurak, K. Miyata, M. Fischer, M. Balakrishnan, S. Chhajed,D. Wells, Li Hong, A. Torsi, J. Lim, M. Korber, K. Nakazawa, S. Mayuzumi, M. Honda, S. Sills,S. Yasuda, A. Calderoni, B. Cook, G. Damarla, H. Tran, Bei Wang, C. Cardon, K. Karda, J. Okuno, A. Johnson, T. Kunihiro, J. Sumino, M. Tsukamoto, K. Aratani, N. Ramaswamy, W. Otsuka,K. Prall. "Process integration of a 27nm, 16Gb Cu ReRAM", Electron Devices Meeting (IEDM), pp.6.2.1-6.2.4, 2014.

[9] T.Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap et al., “A 130.7-mm2 2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology”, IEEE J. Solid-State Circuits 49, pp. 140-153, 2014.

[10] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, et al., “An 8 Mb Multi-Layered Cross-Point ReRAM Macro With 443 MB/s Write Throughput”, IEEE J. Solid-State Circuits 48, pp. 178-185, 2013.

[11] H.-S P. Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. Chen, and M.-J. Tsai, “Metal–Oxide RRAM”, Proceedings of IEEE, vol. 100, pp. 1951-1970, 2012, doi: 10.1109/JPROC.2012.2190369

[12] S. Aldana, P. García-Fernández, A. Rodríguez-Fernández, R. Romero-Zaliz, M.B. González, F. Jiménez-Molinos, F. Campabadal, F. Gómez-Campos, J.B. Roldán, "A 3D Kinetic Monte Carlo simulation study of Resistive Switching processes in Ni/HfO2/Si-n+-based RRAMs", Journal of Physics D: Applied Physics, 50, 335103, 2017.

[13] J. Guy, G. Molas, P. Blaise, M. Bernard, A. Roule, G. Le Carval, V. Delaye, A. Toffoli, G. Ghibaudo, Fellow, IEEE, F. Clermidy, B. De Salvo, L. Perniola , “Investigation of Forming, SET, and Data Retention of Conductive-Bridge Random-Access Memory for Stack Optimization”, IEEE Transactions on Electron Devices, 62, pp. 3482-3489, 2015

[14] A. Padovani, Member, IEEE, L. Larcher, Member, IEEE, O. Pirrotta, L. Vandelli, G. Bersuker, Member, IEEE, “Microscopic Modeling of HfO x RRAM Operations: From Forming to Switching”, IEEE Transactions on Electron Devices, 62(6), pp. 1998-2006, 2015.

[15] M. Bocquet, D. Deleruyelle, C. Muller, J.M. Portal, “Self-consistent physical modeling of set/reset operations in unipolar resistive-switching memories”. Applied Physics Letters, 98, p. 263507, 2011.

[16] M.A. Villena, J.B. Roldán, M.B. González, P. González-Rodelas, F. Jiménez-Molinos, F. Campabadal, D. Barrera, "A new parameter to characterize the charge transport regime in Ni/HfO2/Si-n+-based RRAMs", Solid State Electronics, vol. 118, pp. 56-60, 2016.

[17] M.A. Villena, M.B. González, J.B. Roldán, F. Campabadal, F. Jiménez-Molinos, F.M. Gómez-Campos y J. Suñé, "An in-depth study of thermal effects in reset transitions in HfO2 based RRAMs", Solid State Electronics, 111, pp. 47-51, 2015.

[18] J.B. Roldán, E. Miranda, G. González-Cordero, P. García-Fernández, R. Romero-Zaliz, P. González-Rodelas, A. M. Aguilera, M.B. González, F. Jiménez-Molinos, “Multivariate analysis and extraction of parameters in resistive RAMs using the Quantum Point Contact model”, Journal of Applied Physics, 123, 014501, 2018.

[19] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan, and H.-S. P. Wong, “Verilog-A Compact Model for Oxide-based Resistive Random Access Memory (RRAM),” International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), pp. 41 – 44, 2014. DOI: 10.1109/SISPAD.2014.6931558

[20] Huang, X. Y. Liu, B. Chen, H. T. Li, Y. J. Wang, Y. X. Deng, K. L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang, and J. F. Kang, “A Physics-Based Compact Model of Metal-Oxide-

Based RRAM DC and AC Operations," IEEE Trans. Electron Devices, vol. 60, no. 12, pp. 4090–4097, Dec. 2013.

[21] G. González-Cordero, F. Jiménez-Molinos, J.B. Roldán, M.B. González and F. Campabadal, "In-depth study of the physics behind resistive switching in TiN/Ti/HfO2/W structures", Journal of Vacuum Science & Technology B 35, 01A110, 2017.

[22] G. Gonzalez-Cordero, J. B. Roldan, and F. Jimenez-Molinos, "Simulation of RRAM memory circuits, a Verilog-A compact modeling approach," in 2016 Conference on Design of Circuits and Integrated Systems (DCIS), pp. 1–6, 2016.

[23] F.M. Puglisi, N. Zagni, L. Larcher, P. Pavan, "Random Telegraph Noise in Resistive Random Access Memories: Compact Modeling and Advanced Circuit Design", IEEE Transactions on Electron Devices, 65, pp. 2964 – 2972, 2018.

[24] F.M. Puglisi, L. Larcher, A. Padovani and P. Pavan, "A Complete Statistical Investigation of RTN in HfO2-Based RRAM in High Resistive State", IEEE Transactions on Electron Devices, 62, pp. 2606-2613, 2015.

[25] G. González-Cordero, M. B. González, F. Jiménez-Molinos, F. Campabadal, and J. B. Roldán, "New method to analyze random telegraph signals in resistive random access memories", Journal of Vacuum Science & Technology B, vol. 37, no. 1, p. 012203, 2019.

[26] D. Ielmini, F. Nardi, and C. Cagli, "Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories," Appl. Phys. Lett., vol. 96, pp. 053503-1–053503-3, 2010.

[27] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Statistical fluctuations in HfOx resistive-switching memory: Part II—Random telegraph noise," IEEE Trans. Electron Devices, vol. 61, no. 8, pp. 2920–2927, Aug. 2014.

[28] G. González-Cordero et al., “Analysis of resistive switching processes in TiN/Ti/HfO2/W devices to mimic electronic synapses in neuromorphic circuits”, Solid-State Electronics, vol. 157, pp. 25–33, 2019.

[29] D. Arumí, S. Manich, R. Rodríguez-Montañés and M. Pehl, "RRAM based random bit generation for hardware security applications," 2016 Conference on Design of Circuits and Integrated Systems (DCIS), Granada, 2016, pp. 1-6. doi: 10.1109/DCIS.2016.7845382

[30] Z. Wei et al., "True random number generator using current difference based on a fractional stochastic model in 40-nm embedded ReRAM," 2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, pp. 4.8.1-4.8.4, 2016. doi: 10.1109/IEDM.2016.7838349

[31] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai and Y. Hayashi, "New analysis methods for comprehensive understanding of Random Telegraph Noise", IEEE International Electron Devices Meeting (IEDM), Baltimore, MD, pp. 1-4, 2009. doi: 10.1109/IEDM.2009.5424230

[32] J. Martin-Martinez, J. Diaz, R. Rodriguez, M. Nafria and X. Aymerich, "New Weighted Time Lag Method for the Analysis of Random Telegraph Signals", IEEE Electron Device Letters, vol. 35, pp. 479-481, 2014. doi: 10.1109/LED.2014.2304673

[33] J. Martin-Martinez et al., “Characterization of random telegraph noise and its impact on reliability of SRAM sense amplifiers,” in Proc. 5th Eur. Workshop CMOS Variability (VARI), pp. 1–6, 2014.

[34] C. Márquez, N. Rodríguez, F. Gámiz, R. Ruiz, and A. Ohata, “Electrical characterization of Random Telegraph Noise in Fully-Depleted Silicon-On-Insulator MOSFETs under extended temperature range and back-bias operation,” Solid State Electron., vol. 117, pp. 60–65, 2016.

[35] G. González-Cordero, M. B. González, F. Campabadal, F. Jiménez-Molinos, and J. B. Roldán, “A new technique to analyze RTN signals in resistive memories,” Microelectronic Engineering, vol. 215, p. 110994, 2019.

[36] M. B. Gonzalez, J. Martin-Martinez, M. Maestro, M. C. Acero, M. Nafría and F. Campabadal, "Investigation of Filamentary Current Fluctuations Features in the High-Resistance State of Ni/HfO2-Based RRAM," in IEEE Transactions on Electron Devices, vol. 63, no. 8, pp. 3116-3122, Aug. 2016.

[37] T. Kohonen, “Self-organized formation of topologically correct feature maps,” Biological Cybernetics, vol. 43, no. 1, pp. 59–69, 1982.

[38] T. Kohonen, Self-Organizing Maps, vol. 30. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001.

[39] S. Lloyd, "Least squares quantization in PCM," IEEE Transactions on Information Theory, 28, pp. 129-137, 1982.

[40] M. B. Gonzalez, J. M. Rafí, O. Beldarrain, M. Zabala and F. Campabadal, "Analysis of the Switching Variability in Ni/HfO2 -Based RRAM Devices," in IEEE Transactions on Device and Materials Reliability, vol. 14, no. 2, pp. 769-771, June 2014. doi: 10.1109/TDMR.2014.2311231

[41] M.B. Gonzalez, J. Martin-Martinez, R. Rodriguez, M.C. Acero, M. Nafria, F. Campabadal, X. Aymerich, Dedicated random telegraph noise characterization of Ni/HfO2-based RRAM devices, Microelectronic Engineering, Volume 147, pp. 59-62, 2015, http://dx.doi.org/10.1016/j.mee.2015.04.046.

[42] M. Maestro, J. Martin-Martinez, J. Diaz, A. Crespo-Yepes, M.B. Gonzalez, R. Rodriguez, F. Campabadal, M. Nafria, X. Aymerich, Analysis of Set and Reset mechanisms in Ni/HfO2-based RRAM with fast ramped voltages, Microelectronic

Engineering, volume 147, pp. 176-179, 2015. http://dx.doi.org/10.1016/j.mee.2015.04.057.

[43] F. Puglisi, P. Pavan, "Factorial hidden Markov model analysis of Random Telegraph Noise in Resistive Random Access Memories", in ECTI Transactions on Electrical Eng., Electronics and Communications, vol. 12, no. 1, pp. 24-29, 2014.

[44] R. Carboni, D. Ielmini, “Stochastic Memory Devices for Security and Computing”, Adv. Electron. Mater., vol. 5, Iss. 9, 1900198, 2019.

[45] T. J. Sejnowski, “The Deep Learning Revolution”, MIT Press, 2018.

[46] Y. LeCun, Y. Bengio, G. Hinton. “Deep learning”. Nature. Vol 521, pp. 436-444, 2015 . doi: 10.1038/nature14539.

[47] G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks”, Science, vol. 313, pp. 504–507, 2006.

[48] D. P. Kingma, M. Welling, “Auto-Encoding Variational Bayes”. Proceedings of the 2nd International Conference on Learning Representations (ICLR), 2014.

[49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio “Generative Adversarial Networks”, Proceedings of Advances in neural information processing systems, 2014.

[50] C. S. Wicramasinghe, K. Amarasinghe and M. Manic, "Deep Self-Organizing Maps for Unsupervised Image Classification", IEEE Transactions on Industrial Informatics, p. 1, 2019. doi: 10.1109/TII.2019.2906083