CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Application of artificial neural networks to the identification of weak electrical regions in large area MIM structures
Authors	Muñoz-Gorriz, J.;Monaghan, Scott;Cherkaoui, Karim;Suñé, Jordi;Hurley, Paul K.;Miranda, Enrique
Publication date	2021-10-11
Original Citation	Muñoz-Gorriz, J., Monaghan, S., Cherkaoui, K., Suñé, J., Hurley, P. K. and Miranda, E. (2021) 'Application of artificial neural networks to the identification of weak electrical regions in large area MIM structures', Microelectronics Reliability, 126, 114312 (7pp). doi: 10.1016/j.microrel.2021.114312
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1016/j.microrel.2021.114312
Rights	© 2021, Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC BY-NC-ND 4.0 license. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-07 10:33:37
Item downloaded from	https://hdl.handle.net/10468/13388

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Application of artificial neural networks to the identification of weak electrical regions in large area MIM structures

J. Muñoz-Gorriz[a,*], S. Monaghan[b], K. Cherkaoui[b], J. Suñé[a], P. K. Hurley[b], E. Miranda[a]

[a] *Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Cerdanyola del Valles, Spain*
[b] *Tyndall National Institute, University College Cork, Cork, Ireland*

Abstract

Large area metal-insulator-metal (MIM) structures are prone to exhibit weak electrical regions when they are subjected to severe stress conditions. Although the root cause of this problem is hard to identify, it has been attributed to non-uniform oxide thickness, variable dielectric permittivity, correlated defect generation, and fringe effects in capacitors. In this paper, we explore the application of artificial neural networks (ANNs) to the spatial localization of such weak regions. To this end, HfO_2-based MIM structures were electrically stressed with the objective of generating a large number of breakdown spots. These spots are statistically distributed both in size and location over the device area. Two-input/two-output ANNs with different number of neurons and hidden layers were assessed with the purpose of identifying the best and simplest option for detecting where the most severe damage occurs. The obtained results are compatible with previous studies based on spatial statistics techniques. The method can be applied to other systems that exhibit multiple localized failure events.

1. Introduction

The dielectric breakdown (BD) of the oxide layer in metal-insulator-metal (MIM) and metal-insulator-semiconductor (MIS) structures is a serious reliability problem for advanced microelectronics technologies [1]. As is well known, oxide BD is the consequence of the accumulation of traps or defects within the insulator generated by some sort of electrical stress (constant, ramped or arbitrary voltage/current signals). The phenomenon can occur during device operation or can be purposely induced for technology benchmarking. The formation of a percolation path in between the metal electrodes generates a short and, in some cases, a huge release of energy that can cause visible effects on the devices beyond the loss of the insulating capability of the dielectric film. The use of large area capacitors allows investigating the generation of multiple BD events and therefore gaining insight into statistical features related to the localization of the damage. The analysis can help to detect and investigate, if they exist, electrically weak regions in the device under study.

In previous studies [2]–[4], we demonstrated that the methods of spatial statistics are appropriate for assessing the distribution of BD spots in large area MIM devices ($> 10^3$ μm^2). In our case, when a BD event occurs, it leaves a mark on the top metal electrode because of the high thermal effects developed (**Fig. 1**). For multiple BD events, these marks can be mathematically treated as a point pattern enclosed by an observation window (in general the device area). Using specific spatial statistics functions such as the Ripley's *K* or its derivative, namely the pair correlation function (*g*), among others, the distribution of BD events can be examined to elucidate whether there is consistency or not with a complete spatial randomness (CSR) process [5], [6]. Deviation from CSR is an indication of a possible oxide thickness non-uniformity, local variation of the permittivity value, correlated defect generation or the occurrence of edge effects associated with the dissimilar size of the capacitor plates.

In this work, we make a new advance in this area by focusing the attention on artificial neural networks (ANNs) applied to the analysis of the spatial location of multiple BD spots. In particular, we investigate the use of two-input/two-output feedforward multilayer perceptrons (MLPs) as binary classifiers [7]. The inputs are the location (*X*, *Y*) of the BD spots within the device area and the outputs their probable size (Small, Large). This simple yet encouraging exercise shows the power of ANNs in the area of oxide reliability spatial analysis.

* Corresponding author. Jordi.munoz.gorriz@uab.cat
Tel: +34 (93) 511 3530

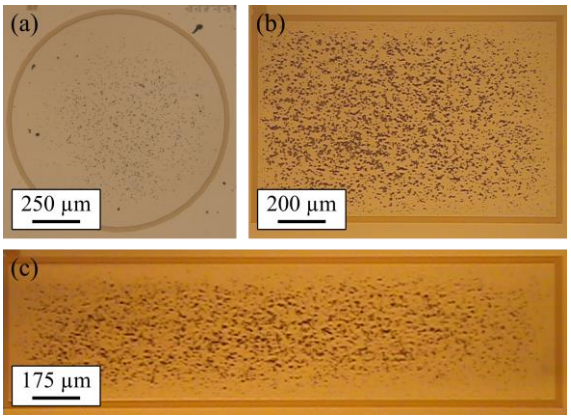


Fig. 1. Photograph of multiple BD spots (a) in a circular area capacitor with diameter Ø=1130 µm and in rectangular area capacitors with size (b) 1250 µm × 800 µm and (c) 2000 µm × 500 µm.

2. Devices and experimental setup

In this paper, Pt/HfO_2/Pt devices with a 30 nm-thick oxide and areas larger than 9.85 × 10^3 µm^2 were investigated. The capacitors were fabricated on n-type Si (100) substrate with resistivity of 1-4 Ω × cm following the steps detailed in [4]. The BD spots were generated by applying a constant voltage stress of -10 V for 120 s. The negative bias was applied to the top electrode with the bottom electrode grounded. **Fig. 1** shows typical optical images corresponding to three devices after the application of the electrical stress. Notice that the BD events leave marks with random sizes on the top electrode. Despite devices with different shapes are available, for the sake of simplicity, the entire discussion that follows will deal exclusively with circular area devices (**Fig. 1**a). The spatial location (*X*, *Y*) and area (A_{BD}) of the BD spots were obtained using MATLAB routines for image analysis. This information was assessed using the *spatstat* package for the R language [8]. The neural network analysis (training and testing) was performed using the software indicated in Ref. [9].

3. Spatial statistics analysis

Fig. 2a,b show the location of all the BD spots detected in two circular area capacitors. **Fig. 2**a corresponds to the spots detected on the device shown in **Fig. 1**a (diameter Ø = 1130 µm) and **Fig. 2**b, to the spots generated on a capacitor with Ø = 564 µm (not shown here). In total, 1389 BD spots were identified in **Fig. 2**a, corresponding to an average intensity λ = 1.38 × 10^{-3} spots/µm^2, and 840 BD spots in **Fig. 2**b, with λ = 3.36 × 10^{-3} spots/µm^2. As expected, more

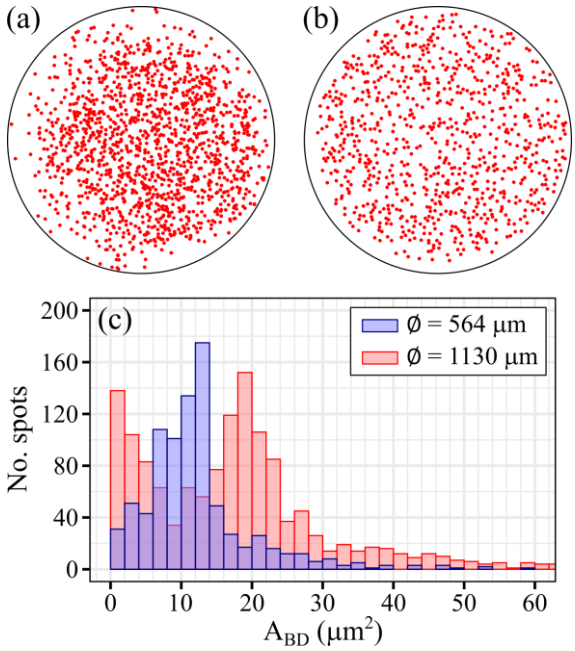


Fig. 2. Location of the BD spots generated in a circular area capacitor with diameter (a) 1130 µm and (b) 564 µm. (c) Histogram for the spot sizes found in (a) and (b).

spots are detected in the device with the largest area [10], [11]. **Fig. 2**c shows a histogram for the size of the spots detected in both devices (**Fig. 2**a,b). From this plot, different distributions can be identified. Whereas for the smallest device, a single peak at 13 µm^2 is detected, for the largest one, two different peaks stand out: one for the small sizes (0 – 5 µm^2), and a second for the large sizes (15 – 25 µm^2). These two peaks will be used to establish a separation of the spots in two groups.

In order to have a clear picture of the BD spots distribution from the perspective of spatial statistics, a number of estimators were computed [6]: the intensity plot (**Fig. 3**a,b), which is a graphical representation of the local density of points; the pair correlation function *g* (**Fig. 3**c,d), which is related to the number of spots located at a distance *r* from each point in the pattern (allows to detect accumulation or repulsion effects); and the mark correlation function K_{mm} (**Fig. 3**e,f), which is a special estimator that evaluates the correlation between the location of the spots with their size. It is important to highlight that just K_{mm} involves the size of the spots whereas the intensity and *g* plots only focus the attention on the location of the spots, as already mentioned. From the intensity plot shown in **Fig. 3**a, and from the experimental value of *g* (**Fig. 3**c), the accumulation of spots towards the center of the device is confirmed for the largest area device (**Fig. 2**a). Notice that this can also be observed in the optical image (**Fig. 1**a).

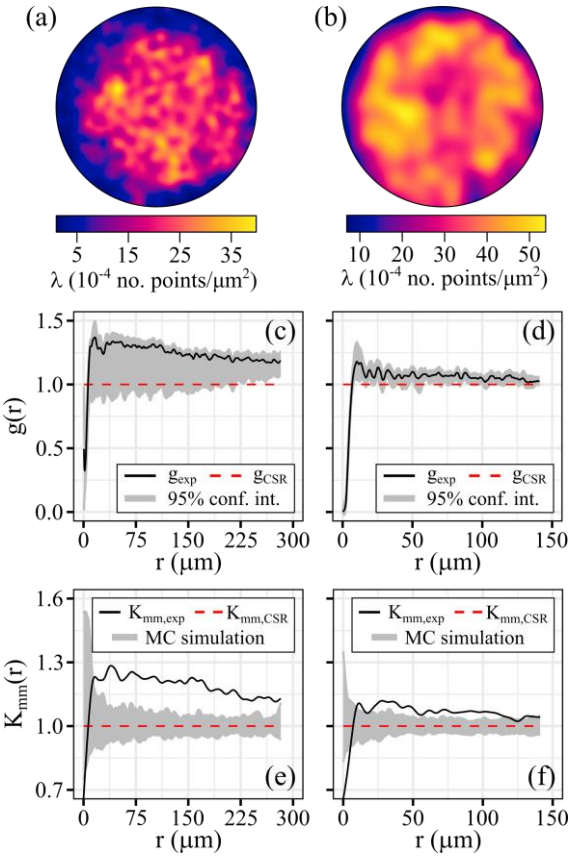


Fig. 3. Estimators for the point pattern shown (a, c, e) in **Fig. 2**a, and (b, d, f) in **Fig. 2**b. (a, b) Intensity plot. (c, d) Pair correlation function *g* for the experimental data and for a CSR process. (e, f) Mark correlation function *K* for the experimental data and for a CSR process. *r* indicates the distance between points in the pattern.

Moreover, the 95% confidence band for the experimental value (shaded region in **Fig. 3**c) includes the value $g = 1$, which indicates that the points likely follow a Poisson process in a reduced observation window within the device area. For the point pattern corresponding to the smallest area device (**Fig. 2**b), the intensity (**Fig. 3**b) and the experimental *g* (**Fig. 3**d) plots show that the location of the spots nearly follows a Poisson distribution. Concerning the correlation between the location and spot size (**Fig. 3**e,f), the shaded region is obtained after 50 Monte Carlo simulations and indicates a remarkable departure from a CSR process. Notice that both patterns show a positive correlation, as the experimental value is higher than $K = 1$, being this correlation stronger for the largest area device (**Fig. 3**e) than for the smallest area one (**Fig. 3**f). This correlation indicates that two spots close to each other tend to have large areas. Moreover, this accumulation of large spots may be interpreted as the existence of electrically weak regions in the device. Identifying where the accumulation of large spots occurs is also possible using the methods of spatial statistics by considering some sort of screening first and then applying the spatial estimators. In this work, we introduce ANNs for this task. As it will be shown next, the proposed method does not require any previous screening.

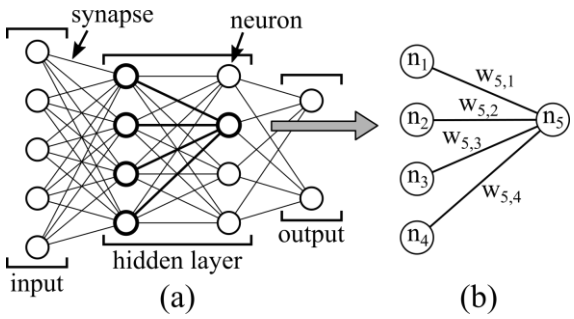


Fig. 4. (a) Schematic representation of a 5:4:4:2 multilayer perceptron network. (b) Representation of the 4 neurons in the first hidden layer and their connections to 1 neuron in the second hidden layer.

4. Artificial neural networks

Before beginning with the analysis, it is mandatory to present a brief explanation on what a neural network or, more specifically, a MLP network is. A MLP is a system formed by neurons and synapses (**Fig. 4**a). The synapses are weights that connect the neurons, and the neurons are functional units distributed in layers. The value of each neuron (**Fig. 4**b), except for the input layer, is determined by the neurons in the previous layer and the values of the attached synapses, that is:

$$n_{out} = F(\boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{n}_{in} + b) \qquad (1)$$

where n_{out} is the output value of a specific neuron, $\boldsymbol{n}_{in} = [n_1, n_2, \ldots, n_n]$ the inputs, and $\boldsymbol{w}^{\mathrm{T}} = [w_{out,1}, w_{out,2}, \ldots, w_{out,n}]^{\mathrm{T}}$ the associated weights. b is the bias (a constant) [12]. Notice that the argument of F in **Eq. (1)** defines in general a kind of hyperplane in the inputs space. F is a nonlinear function, in our case a bipolar sigmoid of the type:

$$F(C) = \frac{2}{(1 + e^{-\beta \cdot C})} - 1 \qquad (2)$$

which squeezes the combination of inputs, weights, and biases in a range between -1 and 1. In **Eq. (2)**, β determines the slope of the transition and $C = \boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{n}_{in} + b$ (see **Eq. (1)**). In summary, the final outcomes of the MLP network are essentially a succession of

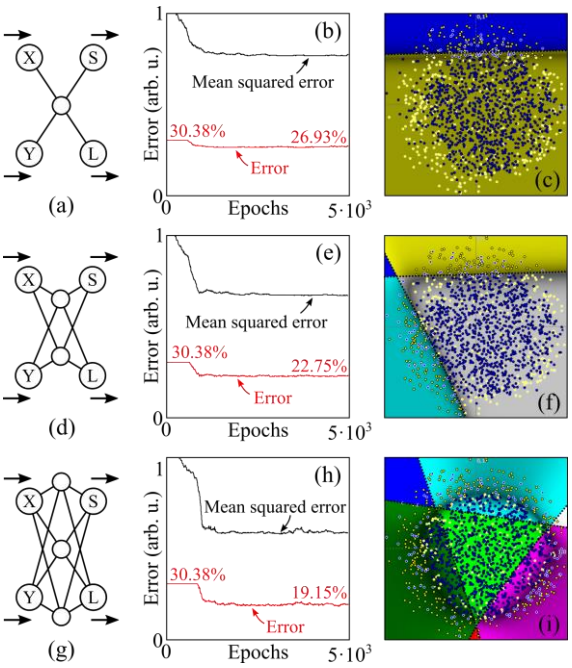


Fig. 5. Analysis of the BD spot distribution shown in **Fig. 2**a with a (a-c) 2:1:2, (d-f) 2:2:2, (g-i) 2:3:2 artificial neural network. The results are obtained after 5×10^3 epochs with a learning rate 0.25. (a, d, g) Schematic representation of the artificial neural network and (b, e, h) its training. (c, f, i) Representation of the classification (decision boundaries) carried out by the artificial neural network.

combinations of nonlinear equations driven by weights and biases. By appropriately selecting weights and biases, the MLP network is able to classify the inputs into different categories. During the learning phase, the output of the ANN corresponding to a given input is compared with the correct output and the weight of the synapses recalculated so as to reduce the quadratic error between the prediction and the desired output. This process is known as supervised learning.

In our case, there are two inputs, both ranging between -1 and 1, corresponding to the *X* and *Y* location of the spot, and two outputs, representing its size (Small or Large). Taking into account the results shown in **Fig. 2**c for the device with Ø = 1130 µm, we separate the points using a threshold value of 10 µm^2. The study that follows is performed on this particular device. The distribution of spots to be analyzed is that shown in **Fig. 2**a.

5. Discussion of results

According to the criterion established (Small < 10 µm^2 < Large) for the selected device (see **Fig. 1**a), the ratio between the small and large spots is 422/967.

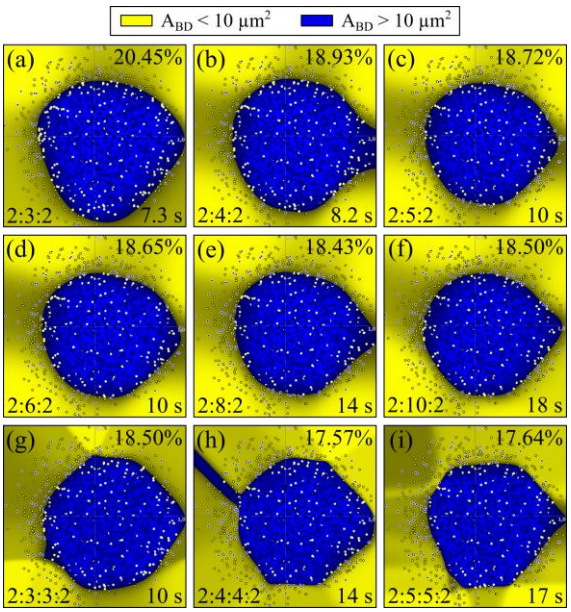


Fig. 6. Outcome of a (a) 2:3:2, (b) 2:4:2, (c) 2:5:2, (d) 2:6:2, (e) 2:8:2, (f) 2:10:2, (g) 2:3:3:2, (h) 2:4:4:2 and (i) 2:5:5:2 artificial neural network after 5×10^3 epochs with a learning velocity rate 0.25 for the point pattern shown in **Fig. 2**a.

Fig. 5 shows the learning process for different ANNs (**Fig. 5**b,e,h) and the computed decision boundaries (**Fig. 5**c,f,i) (2D hyperplanes, *i.e.* lines). Notice that each line splits the pattern into two regions, classifying the points as large or small. The combination of these regions generates the final classification of the spots. In **Fig. 5**c, only one decision boundary is generated. In **Fig. 5**f, two decision boundaries are calculated, which gives a total of four regions. In each of these regions, the classification corresponding to each decision boundary is combined to provide the final answer. Basically, if both decision boundaries classify a region as large, the neural network will recognize that region as large; if the classification of both boundaries is opposite, the stronger one will determine the final outcome of the neural network. As expected, the complexity of the problem increases with the number of decision boundaries. At least three neurons (**Fig. 5**g-i) are required to correctly deal with a possible aggrupation of points. It is important to point out that the error reported in **Fig. 5**b,e,h (red line) refers to the percentage of misclassified points at the beginning and ending of the identification phase. As can be seen, the more complex the artificial neural network is, the lowest the final error is.

Fig. 6 illustrates the final outcome for different MLP networks obtained after 5×10^3 epochs (learning cycles). Notice how all the networks exhibit a similar

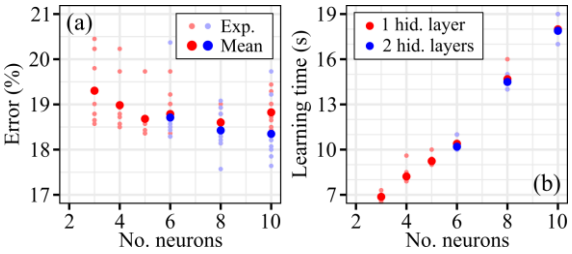


Fig. 7. Representation of (a) the final error and (b) the training time for different neural networks including one and two hidden layers.

trend: a region with large spots is detected close to the center of the point pattern (blue region) surrounded by small spots (yellow region). This confirms not only the positive correlation exhibited by K_{mm} (**Fig. 3**e) but also indicates a clear accumulation of large spots towards the center of the structure, pointing out this region as the most damaged. The MLP outcome is consistent with the results obtained from the spatial statistics methods (**Fig. 3**a,c,e). Importantly, the new method also indicates where the electrical weak region of the device occurs (in terms of the size of spots). Notice that, as the network incorporates more neurons, the shape of the blue region becomes in some cases more irregular and less circular. This is a consequence of the incorporation of more decision boundaries, which in turn generates more regions where the spots can be classified as large. Notice that the point pattern we are dealing with is not completely separable, *i.e.*, there is no region with a single type of spot.

In order to correctly assess the error and learning time, we proceeded several times with the same ANNs shown in **Fig. 6**. In **Fig. 7**a, the mean value and dispersion for the final error corresponding to 10 tests are illustrated. Similarly, the mean value and dispersion for the learning times are illustrated in **Fig. 7**b. As can be seen (**Fig. 7**a), there is a considerable variability in the final error, and the reduction in the mean value is higher from 3 to 5 neurons than from 6 to 10. Moreover, the mean error is somewhat smaller when two hidden layers are considered. On the other hand, for the learning time (**Fig. 7**b), a pretty small dispersion is observed and as expected, the mean value increases with the number of neurons. In this case, no differences are detected between one and two hidden layers. Therefore, for the detection of the accumulation of large spots, simple neural networks with one hidden layer are suitable.

So far, only the point pattern generated on the capacitor with Ø = 1130 µm (**Fig. 2**a) has been analyzed. Despite the distribution of the spots on the device with Ø = 564 µm seems to be consistent with a CSR process and that the positive correlation

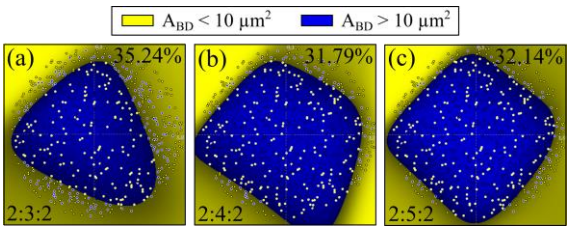


Fig. 8. Outcome of a (a) 2:3:2, (b) 2:4:2 and (c) 2:5:2 artificial neural network after 5 × 10^3 epochs with a learning velocity rate 0.25 for the point pattern shown in **Fig. 2**b.

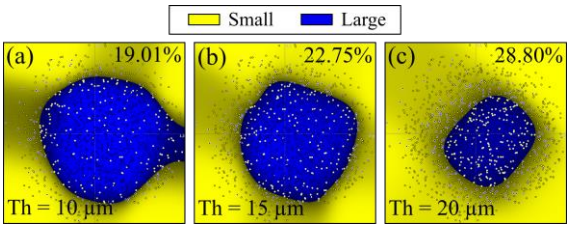


Fig. 9. Outcome of a 2:5:2 artificial neural network after 5 × 10^3 epochs (optimization cycle) with a learning rate 0.25 for the point pattern shown in **Fig. 2**a. The spots are classified as a function of their size: if they are larger or smaller than (a) 10 µm², (b) 15 µm² or (c) 20 µm². According to these criteria, the ratio small/large spots is (a) 422/967, (b) 582/807 and (c) 889/500.

observed in this device is weak, the application of the ANN is still revealing (**Fig. 8**). Recall that each estimator and method in spatial statistics only evaluates a concrete aspect of a point pattern. Thus, several methods are needed to infer the underlying model that governs the distribution of points [2], [5], [13]. In **Fig. 8**, the final outcomes for different MLP networks for the points generated in the smallest area device (**Fig. 2**b) are illustrated. Considering the same criterion for the spots (Small < 10 µm² < Large), in this case, the ratio between the small and large is 334/506. As can be seen, an accumulation of large spots (blue region) towards the center of the device is detected by the neural network. Notice that in some cases, the blue region practically covers the whole device area. This agrees with the weak positive correlation previously observed in **Fig. 3**f.

Finally, it is interesting to assess how the final outcomes of the ANNs change when the classification criterion is modified. So far, the classification was carried out using a threshold value of 10 µm². As the accumulation is more clearly observed in the distribution of BD events generated on the device with the largest area (**Fig. 2**a), the impact of the threshold value is evaluated for this point pattern (**Fig. 9**). In all the cases, a 2:5:2 ANN and 5 × 10^3 epochs were used for benchmarking. Notice that as the threshold value is increased, the blue region reduces indicating where

the accumulation of the largest spots occurs. However, as the number of spots reduces, the total misclassification error increases.

4. Conclusions

In this paper, we demonstrated the suitability of ANNs for investigating the correlation between the location and magnitude of the damage (size of BD spots) induced by a severe electrical stress of MIM structures. The analysis confirms previous observations carried out using the methods of spatial statistics. This points out the compatibility of both methods. ANNs with simple layouts, *i.e.* with one hidden layer and a few neurons are sufficient for correctly classifying the size of the failure events. For the investigated devices, the most damaged region (electrically weak) is found to be close to the center of the device. MLP analysis with an increased number of outputs is proposed to elucidate more in detail if the size of the spots has a more specific connection with their locations.

Acknowledgements

This work was supported by the Spanish Ministry of Science and Universities and the FEDER program through project TEC2017-84321-C4-4-R. The authors also acknowledge the financial support of SFI through the AMBER 2 project (12/RC/2278-P2).

References

[1] A. S. Oates, "Reliability issues for high-k gate dielectrics," in *IEEE International Electron Devices Meeting 2003*, pp. 923–926, 2003.

[2] J. Muñoz-Gorriz *et al.*, "Detection of inhibitory effects in the generation of breakdown spots in HfO_2-based MIM devices," *Microelectron. Eng.*, vol. 215, pp. 111023, 2019.

[3] X. Saura *et al.*, "Analysis of the breakdown spot spatial distribution in Pt/HfO_2/Pt capacitors using nearest neighbor statistics," *J. Appl. Phys.*, vol. 114, no. 15, p. 154112, 2013

[4] J. Muñoz-Gorriz *et al.*, "Exploratory study and application of the angular wavelet analysis for assessing the spatial distribution of breakdown spots in Pt/HfO_2/Pt structures," *J. Appl. Phys.*, vol. 122, no. 21, p. 215304, 2017.

[5] J. Illian *et al., Statistical analysis and modelling of spatial point patterns*. Chichester, UK: John Wiley & Sons, Ltd, 2008.

[6] A. Baddeley *et al., Spatial point patterns : methodology and applications with R*. New York: Chapman and Hall/CRC, 2015.

[7] A. Géron, *Hands-on Machine Learning with Scikit-Learn , Keras & TensorFlow*. Canada: O'Relly Media, Inc, 2019.

[8] A. Baddeley and R. Turner, "Spatstat : an R package for analyzing spatial point patterns," *J. Stat. Softw.*, vol. 12, no. 6, pp. 1–42, 2005.

[9] P. Chlebek, "Sharky Neural Network." [software], 2009. Available from: http://www.sharktime.com/en_SharkyNeuralNetwork.

[10] E. Y. Wu and J. Suñé, "On voltage acceleration models of time to breakdown—Part I: experimental and analysis methodologies," *IEEE Trans. Electron Devices*, vol. 56, no. 7, pp. 1433–1441, 2009.

[11] F. Palumbo *et al.*, "A review on dielectric breakdown in thin dielectrics: silicon dioxide, high-k, and layered dielectrics," *Adv. Funct. Mater.*, vol. 30, no. 18, p. 1900657, 2020.

[12] I. Goodfellow *et al., Deep Learning*, MIT Press, 2016.

[13] J. Muñoz-Gorriz *et al.*, "Spatial analysis of failure sites in large area MIM capacitors using wavelets," *Microelectron. Eng.*, vol. 178, pp. 10–16, 2017.