Failure Analysis of Large Area Pt/HfO_2/Pt Capacitors Using Multilayer Perceptrons

J. Muñoz-Gorriz[1], S. Monaghan[2], K. Cherkaoui[2], J. Suñé[1], P.K. Hurley[2], E. Miranda[1]
[1]Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain
[2]Tyndall National Institute, University College Cork, Cork, Ireland
Phone : (+34) 93 511 3530, Email : Jordi.munoz.gorriz@uab.cat

***Abstract*—In this work, we investigated the spatial distribution of failure sites in large area Pt/HfO_2/Pt capacitors using simple neural networks as classifiers. When an oxide breakdown (BD) happens due to severe electrical stress, a mark shows up in the top metal electrode at the location where the failure event took place. The mark is the result of a microexplosion occurring inside the dielectric film. Large area devices need to be studied because the number of spots required must be sufficient for statistical analysis. The obtained results using multilayer perceptrons with different number of neurons and hidden layers indicate that the largest breakdown spots tend to concentrate towards the center of the device. This observation is consistent with previous exploratory analysis using spatial statistics techniques. This exercise shows the suitability of multilayer perceptrons for investigating the distribution of failure sites or defects on a given surface.**

Keywords—neural networks, perceptron, spatial statistics, MIM, dielectric breakdown, reliability

I. Introduction

Failure analysis of metal-insulator-metal (MIM) capacitors usually involves investigating breakdown (BD) data for a large number of small area devices. Very often, a constant voltage or current stress is applied to each structure and the time-to-BD registered. Subsequently, this information is represented using appropriate reliability plots, such as the Weibull plot. However, in general, this approach does not provide insight into the connection between the size of the BD spot and its spatial location on the device area. This could be of interest for investigating for instance the occurrence of possible time and/or spatial correlations among the generated failure events.

In previous studies [1]–[3], we demonstrated that when the BD spots are visible on the top metal electrode thanks to the generation of localized thermal damage, a complementary analysis is feasible. Instead of a large number of small devices, a single large area device with multiple BD spots can be investigated using spatial statistics techniques. Briefly, the location of the BD spots is mathematically considered as a 2D random point pattern enclosed by a given observation window. Different functions (intensity, Ripley's K, pair correlation, etc.) allow investigating whether this point pattern is consistent with a complete spatial randomness (CSR or Poisson-distributed) process or not [4], [5]. Although deviations from CSR are indications that the points are not Poisson-distributed within the observation window, in general, a number of alternative studies must be carried out in order to infer the correct distribution of the BD spots [6]. If the BD spots are not randomly distributed over the device area, this could mean that some hidden variable is affecting the generation process. This could be oxide thickness nonuniformity, local variations of the permittivity, capacitor edge effects, etc. [7], [8].

This work was supported by the Spanish Ministry of Science, Innovation 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).

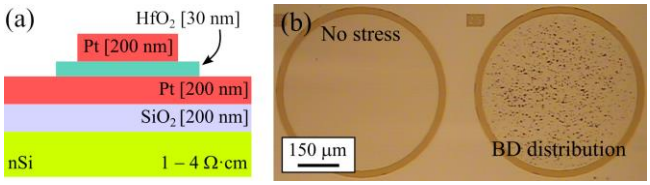

Fig. 1. (a) Cross-sectional sketch of the investigated devices. (b) Optical images of a pristine device and a stressed one.

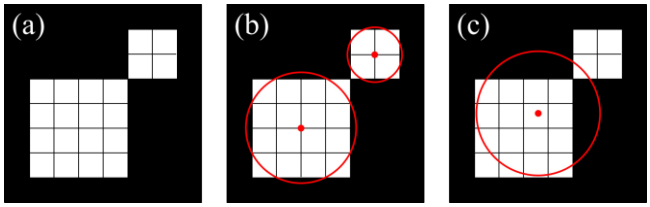

Fig. 2. (a) Schematic representation of a binary image. Spots detected from the anaylisis using (b) 4-connected or (c) 8-connected pixels approach.

The methods of spatial statistics evaluate not only the distribution of the spots but also the possible correlation between the spot locations and their characteristic features, in particular their size. This is interesting as the size of the spots can be used as an additional indicator of the damage induced by the BD phenomenon. In another words, correlation between the location and the size of the spots can be used to detect weak regions of the device. In this work, we go a step further by analysing the failure data (size and location of the BD spots) using simple neural networks. In particular, we focus the attention on the use of multilayer perceptrons as binary classifiers [9]. In this case, the input is the location of the BD spot and the output is its probable size (large or small). Of course, this can be complicated further by considering a more detailed classification scheme. We demonstrate how, after training (synaptic weights determination), the network is able to identify the region where the most severe damage occurred, confirming the results provided by spatial statistics. This is a simple yet encouraging exercise showing the power of neural networks applied to an oxide reliability problem.

II. Devices and Experimental Setup

In this work, MIM structures with a 30 nm-thick HfO_2 dielectric film and Pt electrodes were investigated (see **Fig. 1**a). The capacitors were fabricated on n-type Si (100) substrate with resistivity 1-4 Ω × cm. A detailed description of the fabrication steps can be found in [3]. For the sake of simplicity, a circular area device is exclusively investigated here but the concepts discussed below apply to other device geometries as well. The BD spots were generated by applying a constant voltage stress to the top electrode with the bottom electrode grounded (-10 V for approx. 120s). At the end of the stress, the final state of the device was inspected by means of an optical microscope (see **Fig. 1**b).

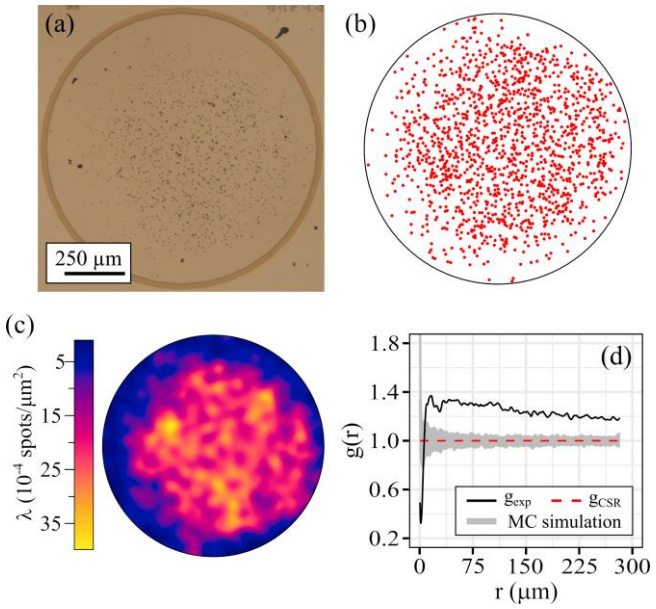

Fig. 3. (a) Photograph of a multiple BD spot distribution in a circular area capacitor with Ø = 1130 µm. (b) Location of points in the capacitor area. (c) Intensity plot for the experimental data. (d) Pair correlation function g for the experimental data and for a CSR process. The shaded region is obtained after 100 Monte Carlo simulations.

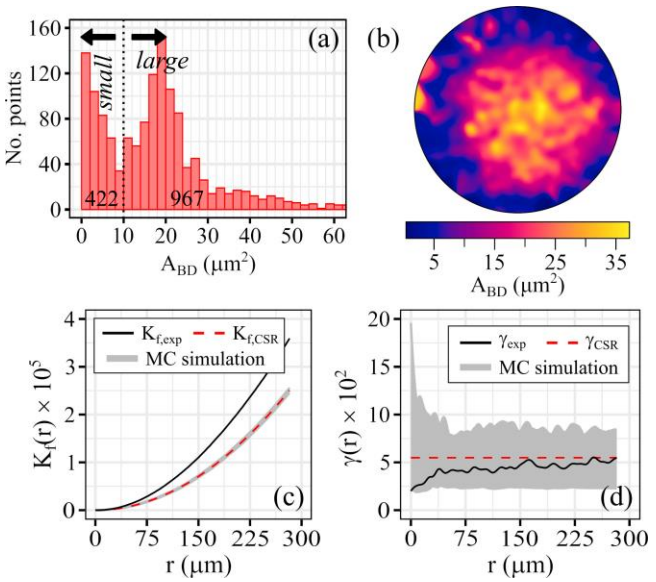

Fig. 4. (a) Histogram for the spot sizes. (b) Smoothing the distribution of BD using the Nadaraya-Watson kernel smoothing, after removing all the spots higher than 100 µm². (c) Mark-weighted K-function K_f and (d) mark variogram γ for the experimental data (**Fig. 3**b) and for a CSR process. The shaded regions are obtained after 100 Monte Carlo simulations.

As can be seen in this figure, the top metal area is plenty of black dots with different sizes. The location and size of each spot is obtained from these optical images using MATLAB routines for image processing. The subsequent process is as follows: first, the image is transformed into a binary image using a threshold value in which the BD spots appear as white regions (**Fig. 2**a). Then, a 4-connected pixel approach (**Fig. 2**b) is used to avoid consider as one two touching spots. In this approach, two pixels are connected if they share an edge. Notice that in an 8-connected pixel approach (**Fig. 2**c), the spots are connected if they also share a corner. Finally, this information is analyzed using the *spatstat* package for the R language [10]. The neural network analysis discussed in Section IV is carried out using the software indicated in [11].

III. Spatial Statisics

A. Distribution of the BD spots

Before considering the size of the BD spots, an evaluation of their spatial distribution is mandatory to correctly interpret the possible correlation between the location of the spots and their size. In **Fig. 3**a, an optical image of the studied device is shown. This device has a large area (Ø = 1130 µm) which allows the generation of a huge number of BD spots [12], [13]. **Fig. 3**b shows the location of all the 1389 BD spots detected in **Fig. 3**a, which correspond to an average intensity $\lambda = 1.38\times10^{-3}$ spots/µm². As the average intensity does not give any clue about the distribution of the spots, the local variation of this parameter is evaluated instead (see **Fig. 3**c). As can be seen, a clear accumulation of spots towards the center of the device is observed. In order to complement this information, the pair correlation function (g) is computed. In **Fig. 3**d, the experimental g and the expected value for a CSR distribution ($g = 1$) are compared. The shaded region represents the expected error for a CSR process and is obtained from 100 Monte Carlo simulations. As can be seen, the accumulation of spots is confirmed as the value of the experimental g is larger than $g = 1$, lying well above the confidence band obtained by the Monte Carlo simulations.

B. Size and location of the BD spots

In Section IIIA, the location of the spots was exclusively evaluated (**Fig. 3**). Now, in this Section, the size of the BD spots will be considered as well. It is worth pointing out that likely even though a 4-connected pixel approach is considered, several BD spots may still merge as a single count. As we are inspecting the final state of the device (**Fig. 3**a), it is not possible to identify if any large spot is in fact the result of several BD events. This does not alter our analysis as large spots still identify regions severely damaged. For the correct assessment of these spots, the time evolution of the damage should be investigated.

Fig. 4a shows a histogram for the area of the BD spots (A_{BD}) detected in **Fig. 3**b. Two clear peaks can be distinguished: one in the range up to 10 µm² and a second one in the range from 10 to 26 µm². These are the most common BD spot sizes. In **Fig. 4**b, the 2D distribution of the Nadaraya-Watson kernel smoothing is shown [5]. Briefly, this plot is similar to the intensity plot, but it takes into account the size of the spots as well, *i.e.* it gives the location of A_{BD}. As can be seen, this plot strongly suggests a concentration of the largest spots towards the center of the device. It is also worth mentioning that **Fig. 4**b corresponds to the distribution shown in **Fig. 3**b after removing all the spots larger than 100 µm². Only 13 spots were removed (0.9 % of the BD spots). **Fig. 4**c and **Fig. 4**d show the mark-weighted K-function (K_f) [14] and the mark variogram (γ) [15] for the distribution shown in **Fig. 3**b, respectively. Considering the accumulation of spots previously demonstrated by the experimental g (**Fig. 3**d), the experimental value of K_f (**Fig. 4**c) indicates that there is a concentration of spots with large A_{BD} as its value is higher than the expected value for a CSR distribution and the confidence band. In addition, the experimental γ (**Fig. 4**d) nearly agrees with a CSR process. This serve as an example why several estimators and methods are needed to evaluate any point pattern as we know from the other estimators that the distribution is not CSR.

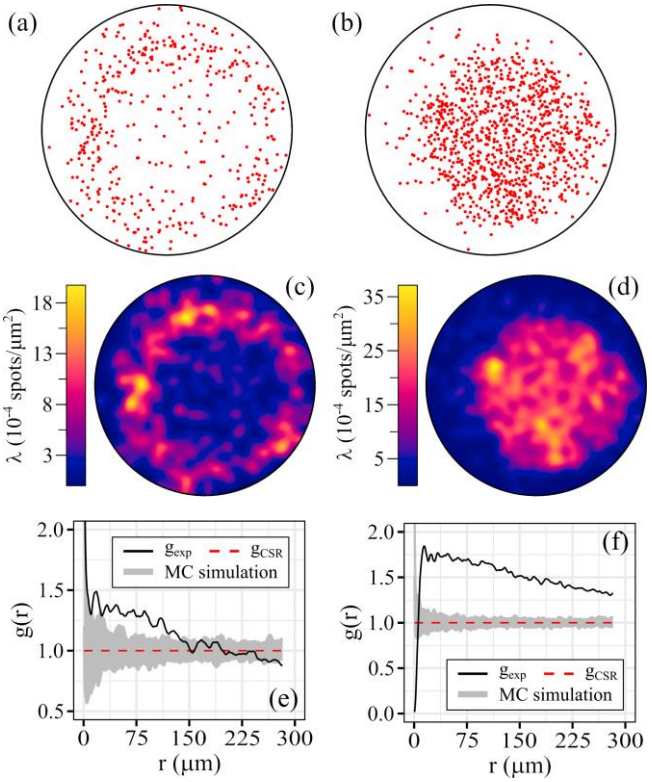

Fig. 5. Location of the BD spots (a) smaller than 10 μm^2 and (b) larger than 10 μm^2. Intensity plot for the distribution of (c) small spots and (d) large spots. Pair correlation function g for the experimental data and for a CSR process corresponding to the distribution of (e) small spots and (f) large spots. The shaded regions are obtained after 100 Monte Carlo simulations.

Since a binary classification will be applied, the points are separated into two groups: Large and Small, with a threshold value 10 μm^2. According to this criterion, there are 422 small spots (**Fig. 5**a) and 967 large spots (**Fig. 5**b). These distributions are assessed using the intensity plot and the estimator g (**Fig. 5**d-f). Interesting, the smallest spots seem to be largely situated near the edge of the device (**Fig. 5**c), which is confirmed by the experimental g value (**Fig. 5**e). As can be seen, its value is higher than the confidence band for short r distances and decreases as the distances increase. On the other hand, notice that the accumulation of the largest BD spots towards the center of the device is evident again: as can be observed in the intensity plot (**Fig. 5**d) and the experimental g value (**Fig. 5**e), the largest spots show accumulation. With this prior knowledge at hand, we will check whether a simple neural network is able to identify the most severe damage region of the device or not.

IV. Multilayer Perceptron Analysis

Briefly, a multilayer perceptron is a system formed by neurons and synapsis (**Fig. 6**). Whereas the synapsis are weights that connects the neurons, and have a fixed value; the neurons are functional units distributed in layers whose value is determined by the expression

$$n_i = F(\boldsymbol{w} \cdot \boldsymbol{n}_j^T + b) \quad (1)$$

where n_i is the value of a specific neuron, $\boldsymbol{n}_j$ is a vector with the value of the previous neurons connected to n_i, $\boldsymbol{w}$ is a vector that contains the associated weights, and b is the bias (a constant) [16]. As an example, the value of n_{10} (see **Fig. 6**) is:

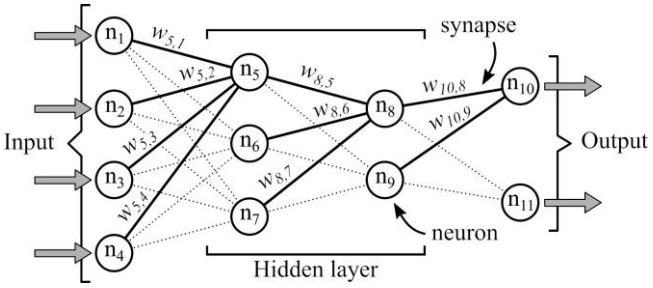

Fig. 6. Schematic representation of a 4:3:2:2 multilayer perceptron network, highlighting the neurons and synapsis that determines the value of some internal neurons.

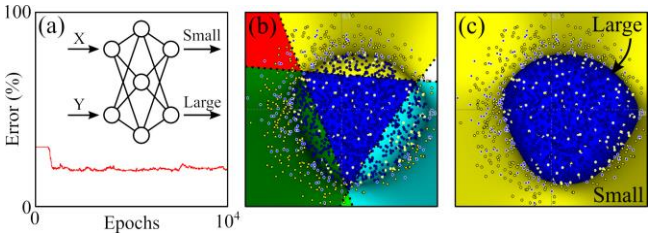

Fig. 7. Analysis of the BD spot distribution shown in Fig. 2b with a 2:3:2 neural network. (a) Training of the neural network and its schematic representation. (b) Representation of the clasification (decision boundaries) carried out by the neural network. (c) Final results provided by the neural network. The blue region corresponds to large spots (A_{BD} > 10 μm^2) and the yellow one to small spots (A_{BD} < 10 μm^2).

$$n_{10} = F([w_{10,8}, w_{10,9}] \cdot [n_8, n_9]^T + b) \quad (2)$$

F is a nonlinear function that squeezes the combinations of neurons, weights, and bias between two values. In our case, these two values are -1 and 1:

$$F(C) = 2 / [1 + \exp(-\beta \cdot C)] - 1 \quad (3)$$

Notice that F is a bipolar sigmoid function where β determines the slope of the transition and C is $\boldsymbol{w} \cdot \boldsymbol{n}_j^T + b$ **(1)**. As the values of the neurons are determined by **(1)**, the learning of the network is performed by modifying the weights of the synapsis. The optimal weights are obtained using a backpropagation learning algorithm, *i.e.*, the desired output is compared with the predictions of the neural network. A minimization of the quadratic error at the output is performed by recalculating the weight of the synapsis. The neural networks used in this paper have a very simple structure, with only two inputs and two outputs. While the inputs represent the X and Y location of a BD spot, the outputs represent the classification previously mentioned: Small or Large. It is important to point out that both coordinates, X and Y, are also squeezed between -1 and 1.

Fig. 7a illustrates the learning process for a multilayer perceptron with one hidden layer and three neurons. Notice that the error reflects the percentage of misclassified spots. At the beginning, the neural network starts classifying all the BD spots as large as there are more large spots than small ones. Then, after some learning cycles, the error drops significantly and starts to fluctuate around a minimum value. This happens when the network is not able to further improve its outcome. The value of each neuron is determined by **(2)**, which defines a kind of hyperplane in the output space. At the end, with the aid of the sigmoid function **(3)**, this 2D hyperplane splits the pattern into two regions, classifying the points in one region

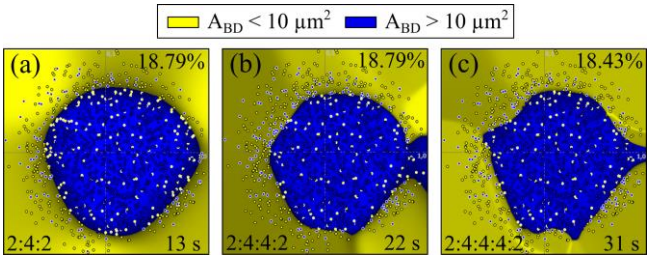

Fig. 8. Outcome of (a) 2:4:2, (b) 2:4:4:2, and (c) 2:4:4:4:2 neural network after 10^4 epochs with a learning velocity of $ni = 0.25$.

as small and the other as large. In **Fig. 7**b, the decision boundary determined by each neuron at the end of the learning process is represented by a dotted line. As can be seen, these three lines discriminate the point pattern in a total of 6 regions. In each region, the classification performed by each hyperplane is combined to determine the final outcome of the network. Basically, if all three decision boundaries classify one of these regions as small, then the neural network will recognize that region as small. Otherwise, if the classification is not the same for all the decision boundaries, both answers (small and large) are evaluated and the dominant one determines the final outcome. In **Fig. 7**c, the final answer of the neural network is illustrated. Notice that the analysis clearly indicates that the large spots (blue region) accumulate towards the center of the device, whereas the small spots (yellow region) are mostly located close to the edge of the structure. This result not only indicates where the most severe damaged region of the device is, but also confirms the results previously observed using the spatial statistics techniques (**Fig. 5**).

In order to evaluate the impact of the selected neural network, more complex networks were analyzed (**Fig. 8**). The results were obtained after 10^4 epochs (optimization cycle). Notice that despite the error reduces as the neuronal network increases its complexity, the shape of the blue region becomes more irregular and less circular. This happens because the distribution of small and large spots are not completely separable, as can be seen when their respective distributions are plotted (**Fig. 5**a,b). As the network becomes more complex, the combination of the decision boundaries leads to more intricate results. In fact, using a very complex network, with a larger number of neurons and several hidden layers, it should be possible to drastically reduce the error as the network would be able to identify several clusters for each type of spot. However, this general classification (**Fig. 8**a) is what we are looking for since we are interested in identifying the most damage region of the device. Therefore, for this type of problem, simple neural networks with one hidden layer do the expected job.

V. CONCLUSIONS

In this paper, the potential of neural networks for investigating oxide reliability issues was demonstrated. Both the methods of spatial statistics as well as the multilayer perceptrons approach indicate that there is an accumulation of more and larger BD spots in the central region of the device. Even though an initial training process is needed, the neural network analysis is a good complementary technique to those of spatial statistics. Although a number of causes have been proposed, the reason behind the accumulation of spots is still unclear so more efforts in investigating this phenomenon are required. The combination of characterization techniques will sure help in that direction.

REFERENCES

[1] 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, p.111023, 2019.

[2] X. Saura, J. Suñé, S. Monaghan, P. K. Hurley, and E. Miranda, "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.

[3] 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.

[4] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical analysis and modelling of spatial point patterns*. Chichester, UK: John Wiley & Sons, Ltd, 2008.

[5] A. Baddeley, E. Rubak, and R. Turner, *Spatial point patterns : methodology and applications with R*. New York: Chapman and Hall/CRC, 2015.

[6] B. D. Ripley, "Modelling spatial patterns," *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, vol. 39, no. 2, pp. 172–212, 1977.

[7] Y. L. Li, Z. Tökei, P. Roussel, G. Groeseneken, and K. Maex, "Layout dependency induced deviation from Poisson area scaling in BEOL dielectric reliability," *Microelectronics Reliability*, vol. 45, no 9-11, pp. 1299–1304, 2005.

[8] X. S. Mas, S. Monaghan, P. K. Hurley, J. Suñé, and E. Miranda, "Failure analysis of MIM and MIS structures using point-to-event distance and angular probability distributions," *IEEE Trans. Device Mater. Reliab.*, vol. 14, no. 4, pp. 1080–1090, 2014.

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

[10] 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.

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

[12] 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.

[13] 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.

[14] A. Penttinen, D. Stoyan, and H. M. Henttonen, "Marked Point Processes in Forest Statistics," *For. Sci.*, vol. 38, no. 4, pp. 806–824, 1992.

[15] N. Cressie, *Statistics for Spatial Data*. New York: JohnWiley & Sons, Inc. 1993.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.