Automatic I–V Parameter Extraction for GaN Devices With Image-Based Machine Learning Method

Zhu, Yi; Schmidt-Szalowski, Marek; Hammes, Petra; Ouhachi, Rezki; Cuoco, Vittorio; Gao, Chang; Tao, Qian; Gajadharsing, John

DOI
10.1109/LMWT.2025.3596254

Publication date
2025

Document Version
Final published version

Published in
IEEE Microwave and Wireless Technology Letters

Citation (APA)
Zhu, Y., Schmidt-Szalowski, M., Hammes, P., Ouhachi, R., Cuoco, V., Gao, C., Tao, Q., & Gajadharsing, J. (2025). Automatic I–V Parameter Extraction for GaN Devices With Image-Based Machine Learning Method. *IEEE Microwave and Wireless Technology Letters*, *35*(12), 2049-2052. https://doi.org/10.1109/LMWT.2025.3596254

Automatic *I–V* Parameter Extraction for GaN Devices With Image-Based Machine Learning Method

Yi Zhu, Marek Schmidt-Szalowski, Petra Hammes, Rezki Ouhachi, Vittorio Cuoco, Chang Gao, *Member, IEEE*, Qian Tao, and John Gajadharsing, *Senior Member, IEEE*

***Abstract*—This study presents a novel image-based machine learning (ML) method for automating *I–V* parameter extraction in gallium nitride (GaN) devices. Using Ampleon's GEAR model, a dataset of 100000 simulated *I–V* curves are converted into *I–V* images through specifically designed transfer functions to train a convolutional neural network. The proposed method outperforms the existing ML method based on a fully connected neural network, particularly for *I–V* curves in the subthreshold region. Validation with measured pulse *I–V* data shows its superior accuracy, achieving a normalized mean square error (NMSE) of −30 dB compared with −24 dB with the existing ML method. The proposed method demonstrates a strong potential to accelerate the extraction and enhance the accuracy of GaN device modeling.**

***Index Terms*—Ampleon GEAR model, convolutional neural network (CNN), device modeling, gallium nitride (GaN) high electron mobility transistor (HEMT), *I–V* parameters, machine learning (ML).**

I. Introduction

Accurate and quick device modeling is essential for circuit simulation and design. In recent years, gallium nitride (GaN) has emerged as a promising semiconductor technology for applications requiring high breakdown voltage and high electron mobility, including radio frequency power amplifiers (RFPAs) [1], [2]. The GaN technology has advanced significantly owing to increasingly faster iteration cycles in process development [3], [4], [5], [6]. Traditional manual methods for extracting model parameters are labor-intensive and time-consuming, failing to keep pace with the rapid evolution of GaN processes. The mismatch between process iteration speed and parameter extraction capability affects the efficiency and performance of model-based circuit design. An efficient and reliable model extraction method is urgently needed.

Recent advancements in machine learning (ML) techniques have demonstrated the potential of neural networks for automated device parameter extraction [7]. One category of approaches replaces the device model, either completely or partially, with a fully connected neural network to represent the complex voltage–current relationship [8], [9], [10], [11], [12]. The other category leverages the existing device model architecture while using ML techniques to accelerate the parameter extraction process. For example, small-signal parameters were extracted using a neural network in [13], [14], and [15] and worked for large-signal parameter extraction [16]. The *I–V* parameters, one of the most important parts in device model, have been proven extractable by neural networks. For example, in [17], [18], [19], [20], [21], and [22], the neural network was used to extract the FinFET device's parameters based on the BSIM model. Similarly, another study has shown that neural networks can automatically extract *I–V* parameters for GaN high electron mobility transistor (HEMT) devices based on ASM-HEMT model [23].

A common feature in these previous works is the fully connected neural network (FCN) architecture, which directly learns the relationship between a series of *I–V* data (input data) and a set of device parameters (output data). This approach, however, differs from the way engineers typically perform the task: in practice, engineers would visualize *I–V* data points as 2-D plots with multiple curves, treating them as images rather than discrete data points. In this process, the focus is on interpreting features of *I–V* images. For instance, specific characteristics such as the slope of the $I_D V_G$ curves are directly associated with the corresponding device parameters in the model. Similarly, comparing the difference between the visualized *I–V* images from the model and the measurements is a very intuitive approach used in the model validation.

These observations inspired our new approach of incorporating 2-D information for automatic *I–V* parameter extraction, in contrast to a 1-D data-to-data mapping. Although neural networks have the capability to learn arbitrarily complex relationships [24], it has been convincingly shown that a proper inductive bias in neural network design, i.e. guiding principles specific to the problem at hand, can solve the problem with better performance, less data, and faster convergence [25]. The most prominent example is the use of convolutional neural networks (CNNs) in solving computer vision problems, benefiting from its deep image prior [26]. Likewise, in this work, we aim to add a domain-specific inductive bias as drawn from the practical experience of modeling engineers, for automatic *I–V* parameter extraction. We show that by mimicking the engineer's effective approach to parameter extraction, the new

Received 19 June 2025; revised 18 July 2025; accepted 24 July 2025. Date of publication 21 August 2025; date of current version 11 December 2025. *(Corresponding author: Yi Zhu.)*

Yi Zhu is with Ampleon B.V., 6534 AV Nijmegen, The Netherlands, and also with the Microelectronics Department, TU Delft, 2628 CD Delft, The Netherlands (e-mail: yi.zhu@ampleon.com).

Marek Schmidt-Szalowski, Petra Hammes, Rezki Ouhachi, Vittorio Cuoco, and John Gajadharsing are with Ampleon B.V., 6534 AV Nijmegen, The Netherlands.

Chang Gao is with the Microelectronics Department, TU Delft, 2628 CD Delft, The Netherlands.

Qian Tao is with the Imaging Physics Department, TU Delft, 2628 CD Delft, The Netherlands.

Digital Object Identifier 10.1109/LMWT.2025.3596254

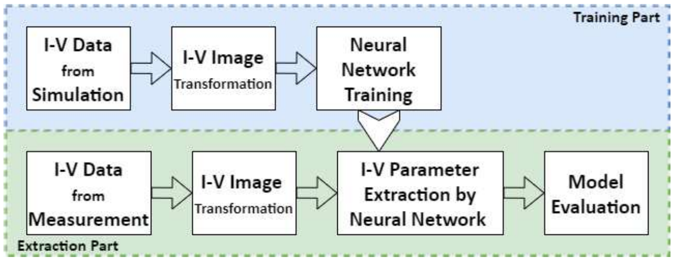

Fig. 1. *I–V* parameter extraction flow with the image-based method.

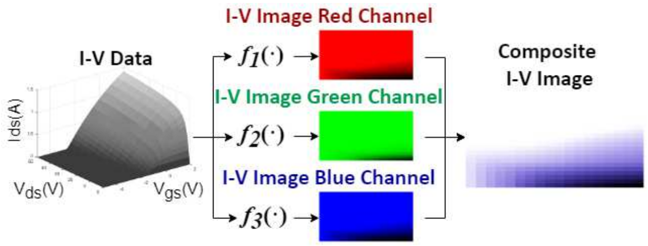

Fig. 2. Conversion of *I–V* data into *I–V* image.

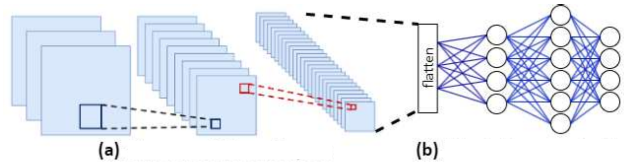

Fig. 3. Structure of a convolution neural network. (a) Convolutional layers. (b) Fully connected layer.

image-based method (mIM) can extract the *I–V* parameter more accurately and efficiently, compared with the existing method based on FCN (mFCN).

II. Parameter Extraction Methodology

The image-based method of *I–V* parameter extraction flow is illustrated in Fig. 1, consisting of two parts: the neural network training process and the *I–V* parameter autoextraction process. The training process begins with generating *I–V* data through simulation, followed by transforming the data into *I–V* images. These images are then used to train a neural network to establish the relationship between *I–V* images and the corresponding model parameters (*I–V* parameters).

Once the neural network is trained, the autoextraction process uses *I–V* data obtained from GaN HEMT measurements. The measured pulse *I–V* data are similarly transformed into *I–V* images and fed into the trained neural network model. The network extracts the *I–V* parameters from these images. Finally, the device model with extracted parameters is evaluated, ensuring the accuracy of extracted parameters.

A. Training Data Generation

To generate sufficient training data for the neural networks, we use Ampleon's GEAR model as the basis for data generation through ADS simulation. The GEAR model is a compact model developed for GaN devices [27]. Ampleon's models and products have been widely used by designers for various applications, including mobile network basestation, broadcast transmitters, medical resonance imaging, etc. [28], [29].

The simulated training data cover the bias conditions within operation range: drain voltage (V_D) $\in$ $[0, 50]$V and gate voltage(V_G) $\in$ $[-5, 2]$V. The GEAR model includes ten *I–V* parameters, each of which is assigned a range wide enough to cover parameter variations based on process statistics. A Monte Carlo simulation is then used to generate 100 000 unique *I–V* parameter combinations within these predefined ranges, ensuring coverage of measurement variations.

B. I–V Image Transformation

In HEMT devices, the drain current I_D is a 2-D function of V_D and V_G, forming a surface over bias space (see Fig. 2). In (1)–(3), x denotes I_D at each (V_D, V_G) point. The transfer functions are applied elementwise: $f_1(x)$ performs global scaling to preserve the overall structure and normalize the range relative to f_2 and f_3; f_2 enhances the subthreshold region using a sigmoid transformation that amplifies low current values, while f_3 emphasizes the saturation region using an exponential transformation that expands high current values. Although the CNN could theoretically model f_1, its explicit inclusion improves numerical conditioning by avoiding channel imbalance and allows the network to focus more effectively on nonlinear patterns captured by f_2 and f_3. This combination improves feature contrast across key regions relevant to parameter learning

$$f_1(x) = \alpha_1 \cdot x + \alpha_2 \tag{1}$$

$$f_2(x) = \beta_1 / \left(1 + e^{-(\beta_2 \cdot x + \beta_3)}\right) \tag{2}$$

$$f_3(x) = \gamma_1^{(\gamma_2 \cdot x + \gamma_3)} \tag{3}$$

where $\alpha_1, \ldots, \gamma_3$ are the coefficients of the transfer functions. These coefficients were empirically chosen based on extensive experience with *I–V* data across GaN devices, aiming to balance output ranges and enhance distinct nonlinear regions. Each transfer function generates a 2-D matrix of transformed current values, corresponding to the original $I_D(V_D, V_G)$ surface. These matrices, similar in structure to image channels used in computer vision, are then mapped to the red, green, and blue channels of a composite *I–V* image, preserving the spatial layout of the data. The resulting image serves as an input to the convolutional neural network. This RGB representation follows standard image processing conventions and also facilitates visual interpretation of device behavior.

C. Neural Network Architecture Selection

After the *I–V* data are transformed through the image transformation method described in Section II-B, a neural network learns from these *I–V* images and their corresponding *I–V* parameters. Among various neural network architectures, convolutional neural network (CNN), initially proposed in [30], features a unique structure suited for learning from images. The CNN architecture, as illustrated in Fig. 3, consists of two parts: convolutional layers and fully connected layers. The convolutional layers [see Fig. 3(a)] extract image features and patterns, such as edges, textures, and shapes. The fully connected layer [see Fig. 3(b)] maps these extracted features to the final output. These layers establish the relationship between the *I–V* image features in the training data and target *I–V* parameters by connecting every neuron in one layer to every neuron in the subsequent layer.

By training a properly designed CNN using the processed *I–V* images, the network can effectively learn the device model parameters from the input images. This approach leverages the ability of CNNs to capture complex, high-dimensional patterns in images, enabling accurate parameter extraction.

III. Results and Discussion

The network used in mIM is designed with 12 layers. The first nine are convolutional layers, each followed by a ReLU activation. These are followed by three fully connected layers with 512, 256, 128, and 10, each paired with a ReLU activation. The final output layer contains ten neurons, corresponding to the target *I–V* parameters.

For comparison, mFCN is implemented as a six-layer fully connected network with hidden layers of 2048, 2048, 1024, 512, 128, and a ten-neuron output layer. Each layer uses ReLU activation, consistent with mIM. To ensure a fair comparison, both the models are trained on the same dataset, with *I–V* data transformed through the same set of transfer functions [see (1)–(3)]. While mIM processes the data as an image, mFCN receives the same *I–V* data flattened into a 1-D vector. Both the models use the Adam optimizer for backpropagation.

The training dataset is generated through simulations of Ampleon's GEAR model, as described in Section II-A, and transformed into images, as detailed in Section II-B. The dataset is divided with 90% for training and 10% for testing. Training is conducted over 500 epochs, as no significant error reduction was observed beyond this point. The process runs on an Intel[1] i7-11850H CPU at 2.5 GHz with an NVIDIA RTX A2000 GPU. mFCN training required approximately 200 min, while mIM took only 75 min.

To evaluate both the methods, the coefficient of determination (R^2) was used [11], [31], which quantifies the correlation between the reference *I–V* parameters and the parameters extracted by the neural networks. These reference parameter sets, numbering 10 000, were randomly distributed within the process variation range and used to generate the corresponding *I–V* data. All the parameter values were then normalized to the interval [−1, 1]. This procedure ensured that the assessment focused on the methods' generalization ability to unseen data, thereby providing a robust comparison of performance.

Fig. 4 shows that for parameters such as GM and GON, both the methods achieve strong correlations. However, for max current parameter IDSAT, mIM significantly outperforms mFCN ($R_{\mathrm{mIM}} > 0.99$ versus $R_{\mathrm{mFCN}} \approx 0.56$). For the subthreshold related parameter such as SS and SS_0, mFCN yields weak correlations ($R_{\mathrm{mFCN}} < 0.5$), while mIM maintains high accuracy ($R_{\mathrm{mIM}} > 0.99$). These results highlight the method's ability to capture fine details, particularly in the subthreshold region, demonstrating strength in accurate parameter extraction.

We evaluated the trained networks using measured pulse *I–V* data from devices fabricated with Ampleon's Gen5 GaN technology. Each device has a unit gate width of 6 × 300 μm. The extracted *I–V* parameters from mIM and mFCN were applied to Ampleon's GEAR model for simulations in ADS. Simulation results and measurements are shown in Fig. 5.

[1]Registered trademark.

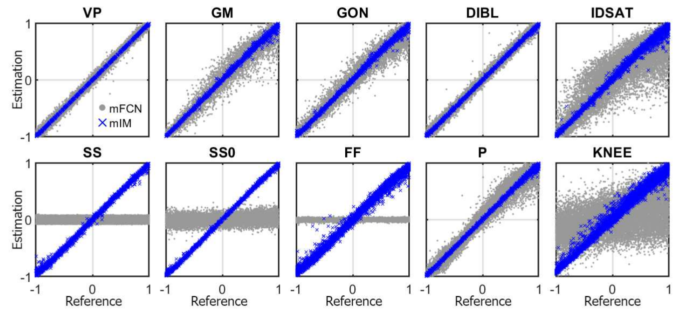

Fig. 4. Comparison of device *I–V* parameters' estimation accuracy between the image-based method (mIM) and the existing method (mFCN).

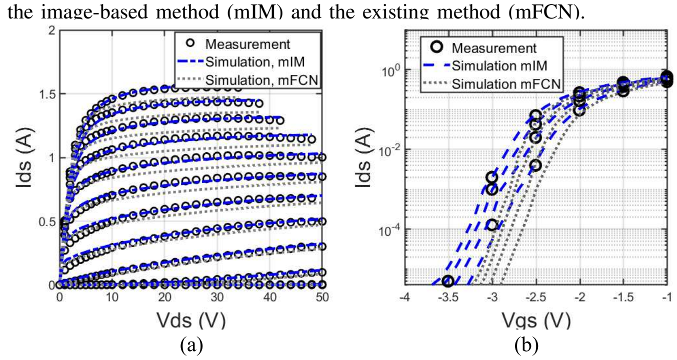

Fig. 5. Comparison of measurement data with simulations using the proposed mIM method and mFCN. (a) $I_D V_D$ characteristics over the full bias range. (b) $I_D V_G$ characteristics highlighting the subthreshold region.

TABLE I

NMSE of Model Simulation Versus Measurements

	Over Full bias	Subthreshhold Region
Model by mIM	-31.7dB	-30.5dB
Model by mFCN	-24.1dB	-19.1dB

In Fig. 5(a), the model with *I–V* parameters generated by both mIM and mFCN shows generally good agreement with the measurements over full bias. However, in terms of accuracy, the model using parameters extracted by the mIM outperformed mFCN. The normalized mean square error (NMSE) for mIM and mFCN is −31.7 and −24.1 dB, respectively (Table I). Fig. 5(b) focuses on the subthreshold region of the $I_D V_G$ curves. In this zoomed-in view, the simulation using the mFCN exhibits notable errors in the slope of I_D. In contrast, the simulation with parameters from the mIM aligns closely with the measurements, providing accurate predictions even in the subthreshold region. The NMSE for mIM in this region is significantly better than that of mFCN, as reflected in Table I. These results align with the correlation statistics shown in Fig. 4, where the image-based method demonstrated strong correlations across all the parameters, including those associated with subthreshold current. This highlights the effectiveness of the image-based method in achieving high accuracy and reliability in device modeling.

IV. Conclusion

This study presents a novel image-based ML method for automating *I–V* parameter extraction in GaN devices. By transforming *I–V* data into images and using convolutional neural networks, the method achieved exceptional accuracy, with correlation determination exceeding 0.99, even in subthreshold regions. The proposed mIM requires less than half of the training time than the conventional mFCN (75 versus 200 min). Validation using measured pulse *I–V* data demonstrated significantly lower NMSE, particularly in subthreshold regions of the *I–V* curves. These results establish that the image-based method efficiently captures complex device characteristics, reducing both computational resources and manual effort required for parameter extraction. Future research could extend the proposed method to other semiconductor technologies, broadening its application scope and enhancing device modeling processes essential for advancing modern semiconductor innovation.

REFERENCES

[1] U. K. Mishra, P. Parikh, and Y.-F. Wu, "AlGaN/GaN HEMTs—An overview of device operation and applications," *Proc. IEEE*, vol. 90, no. 6, pp. 1022–1031, Jun. 2002.

[2] U. K. Mishra, S. Likun, T. E. Kazior, and Y.-F. Wu, "GaN-based RF power devices and amplifiers," *Proc. IEEE*, vol. 96, no. 2, pp. 287–305, Feb. 2008.

[3] R. S. Pengelly, S. M. Wood, J. W. Milligan, S. T. Sheppard, and W. L. Pribble, "A review of GaN on SiC high electron-mobility power transistors and MMICs," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1764–1783, Jun. 2012.

[4] K. J. Chen et al., "GaN-on-Si power technology: Devices and applications," *IEEE Trans. Electron Devices*, vol. 64, no. 3, pp. 779–795, Mar. 2017.

[5] V. Camarchia, R. Quaglia, A. Piacibello, D. P. Nguyen, H. Wang, and A.-V. Pham, "A review of technologies and design techniques of millimeter-wave power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2957–2983, Jul. 2020.

[6] M. Buffolo et al., "Review and outlook on GaN and SiC power devices: Industrial State-of-the-art, applications, and perspectives," *IEEE Trans. Electron Devices*, vol. 71, no. 3, pp. 1344–1355, Mar. 2024.

[7] C. C. McAndrew, A. J. Scholten, K. K. Gullapalli, Y. Chauhan, and K. Xia, "Benchmarks for SPICE modeling and parameter extraction based on AI/ML," *IEEE Trans. Electron Devices*, vol. 72, no. 4, pp. 1551–1559, Apr. 2025.

[8] J. Wang, Y.-H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, "Artificial neural network-based compact modeling methodology for advanced transistors," *IEEE Trans. Electron Devices*, vol. 68, no. 3, pp. 1318–1325, Mar. 2021.

[9] M. Li, O. Irsoy, C. Cardie, and H. G. Xing, "Physics-inspired neural networks for efficient device compact modeling," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 44–49, 2016.

[10] J. Xu and D. E. Root, "Advances in artificial neural network models of active devices," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Aug. 2015, pp. 1–3.

[11] K. Mehta and H.-Y. Wong, "Prediction of FinFET current-voltage and capacitance-voltage curves using machine learning with autoencoder," *IEEE Electron Device Lett.*, vol. 42, no. 2, pp. 136–139, Feb. 2021.

[12] S. Woo, H. Jeong, J. Choi, H. Cho, J.-T. Kong, and S. Kim, "Machine-learning-based compact modeling for sub-3-nm-node emerging transistors," *Electronics*, vol. 11, no. 17, p. 2761, Sep. 2022.

[13] Z. Marinkovi, G. Crupi, A. Caddemi, V. Markovi, and D. M. Schreurs, "A review on the artificial neural network applications for smallsignal modeling of microwave FETs," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 33, May 2020, Art. no. e2668.

[14] F. Chavez and S. Khandelwal, "Deep learning-based ASM-HEMT high frequency parameter extraction," in *Proc. IEEE Wireless Microw. Technol. Conf. (WAMICON)*, Apr. 2023, pp. 41–44.

[15] R. Guan, H. Zhu, W. Che, and Q. Xue, "An improved parameter extraction approach for GaN HEMT small-signal modelling," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 1, Aug. 2022, pp. 1–3.

[16] F. Chavez and S. Khandelwal, "Machine learning-based large-signal parameter extraction for ASM-HEMT," *IEEE Microw. Wireless Technol. Lett.*, vol. 34, no. 2, pp. 147–150, Feb. 2024.

[17] Y. Liu, X. Liu, and P. Cao, "A deep-learning based BSIM-CMG FinFET modeling and parameter extraction approach," in *Proc. Int. Conf. Electron. Eng. Inf. Syst. (EEISS)*, Jan. 2024, pp. 185–188.

[18] F. Chavez, C.-T. Tung, M.-Y. Kao, C. Hu, J.-H. Chen, and S. Khandelwal, "Deep learning-based I–V global parameter extraction for BSIM-CMG," *Solid-State Electron.*, vol. 209, Nov. 2023, Art. no. 108766.

[19] M.-Y. Kao, F. Chavez, S. Khandelwal, and C. Hu, "Deep learning-based BSIM-CMG parameter extraction for 10-nm FinFET," *IEEE Trans. Electron Devices*, vol. 69, no. 8, pp. 4765–4768, Aug. 2022.

[20] A. Singhal, G. Pahwa, and H. Agarwal, "A novel physics aware ANN-based framework for BSIM-CMG model parameter extraction," *IEEE Trans. Electron Devices*, vol. 71, no. 5, pp. 3307–3314, May 2024.

[21] C. Shen et al., "A neural network based fast parameter extraction of compact hot carrier degradation model in FinFETs," in *Proc. 8th IEEE Electron Devices Technol. Manuf. Conf. (EDTM)*, Mar. 2024, pp. 1–3.

[22] F. Chavez, J.-H. Chen, C.-T. Tung, C. Hu, and S. Khandelwal, "I–V global parameter extraction for industry standard FinFET compact model using deep learning," in *Proc. IEEE Int. Symp. Radio-Freq. Integr. Technol. (RFIT)*, Aug. 2023, pp. 20–22.

[23] F. Chavez, D. T. Davis, N. C. Miller, and S. Khandelwal, "Deep learning-based ASM-HEMT I–V parameter extraction," *IEEE Electron Device Lett.*, vol. 43, no. 10, pp. 1633–1636, Oct. 2022.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[25] M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," *Proc. Nat. Acad. Sci. USA*, vol. 116, pp. 15849–15854, Aug. 2019.

[26] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," 2017, *arXiv:1711.10925*.

[27] M. Schmidt-Szalowski, P. Hammes, and V. Cuoco, "From on-wafer measurements to customer models: Ampleon's modeling flow for ldmos and GAN RFPAS," presented at the Eur. Microw. Week (EuMW), Oct. 2017. [Online]. Available: https://www.youtube.com/watch?v=UbPsPsVZt0k

[28] Ampleon.*Simulation Model*. Accessed: Nov. 28, 2024. [Online]. Available: https://www.ampleon.com/support/documentation.html

[29] P. Hammes, M. Nelsy, S. Loysel, V. D. Z. Jos, and M. Schmidt-Szałowski, "A robust, large-signal model for LDMOS RF power transistors," *Microw. RF*, vol. 56, no. 2, p. 41, Jan. 2017.

[30] K. Fukushima, "Visual feature extraction by a multilayered network of analog threshold elements," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-5, no. 4, pp. 322–333, Oct. 1969.

[31] Y. Dodge, "Coefficient of determination," in *The Concise Encyclopedia of Statistics*. New York, NY, USA: Springer, 2008.