POLITECNICO DI TORINO
Repository ISTITUZIONALE

Synergic Exploitation of TCAD and Deep Neural Networks for Nonlinear FinFET Modeling

Original
Synergic Exploitation of TCAD and Deep Neural Networks for Nonlinear FinFET Modeling / Kouhalvandi, L., Catoggio, E., DONATI GUERRIERI, S.. - ELETTRONICO. - (2023), pp. 542-546. (IEEE EUROCON 2023 - 20th International Conference on Smart Technologies Torino, Italy 6-8 July 2023) [10.1109/EUROCON56442.2023.10198982].

Availability:
This version is available at: 11583/2981381 since: 2023-09-08T10:04:10Z

Publisher:
IEEE

Published
DOI:10.1109/EUROCON56442.2023.10198982

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

04 November 2026

Synergic Exploitation of TCAD and Deep Neural Networks for Nonlinear FinFET Modeling

Lida Kouhalvandi[1*,△], Eva Catoggio[2*] , Simona Donati Guerrieri[3*]

[*] *Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy*

[△] *Department of Electrical and Electronics Engineering, Dogus University, Istanbul, Turkey*

lida.kouhalvandi@ieee.org[1], eva.catoggio@polito.it[2], simona.donati@polito.it[3]

***Abstract* — Accurate large-signal (LS) modeling of Fin Field-Effect Transistors (FinFETs) plays an important role in designing microwave circuits for the next generation of communication systems and quantum sensing. In this work we propose preliminary results on a novel approach to LS FinFET modeling, that translates the X-parameters from physical TCAD analysis into deep neural networks (DNNs). The proposed method includes two phases. First, the X-parameters of nonlinear active device are extracted trough accurate TCAD physical simulations. Then, a long short-term memory (LSTM)-based DNN is employed for ANN modelling, to reproduce the scattered waves for any given incident waves up to the 5^{th} harmonic. Similarly to X-parameters, the proposed DNN model simulates the transistor behavior around the large-signal operating point. Unlike the original X-parameter method, though, the DNN approach can incorporate the dependency on bias or other technological and physical parameters in a seamless and numerically efficient way. Hence, once implemented into circuit simulators, it allows for faster and more accurate circuit design.**

***Keywords* — Deep neural network (DNN), fin field-effect transistor (FinFET), long short-term memory (LSTM), large-signal modeling, predict, X-parameter.**

I. Introduction

The need of high-performance semiconductor devices operating in the radio frequency (RF) and microwave frequency range is fostered by many applications ranging from the rapid development of the next generation of communications systems, the precise control and readout of quantum states [1] and the space economy. In this scenario, the FinFET becomes a key enabling technology. In fact, their nanometer scale makes them ideal for quantum sensing while their compatibility with digital applications makes them ideal candidates for the deployment of integrated microwave transceivers in 5G/6G communication networks, e.g. in small-cells and vehicular radars [2].

For these nonlinear circuits, FinFET large-signal models become essential, both at the physical level to optimize the device structure, and at the circuit level for the PA design and optimization [3]–[7]. Accuracy and numerical efficiency are key requirements for nonlinear models, typically addressing the simulation of time-domain waveforms, high-order harmonics, noise and frequency conversion.

Various approaches have been proposed for nonlinear transistor modeling. While compact models based on equivalent circuits are still the most common approach [8], behavioral models are also gaining growing attention, e.g. using polyharmonic distortion [9], X-parameters [10], [11] or the Padé model [12]. Behavioral models are promising solutions for device modeling since they can include more physical features with respect to compact models (e.g. thermal effects, low frequency dispersion, or technological variability) and reproduce characterisation data more accurately. However, their accuracy relies importantly on large characterization data. Furthermore, behavioral models entirely loose the link to the underlying fabrication technology and have poor extrapolation capability. Finally, behavioral models are often based on large look-up tables (LUTs), requiring large memory allocation and repeated numerical interpolation on LUT data.

Recently, intelligent-based methods including artificial neural networks (DNN), have proven their validity for highly accurate FET modeling [13]. By reviewing the fast expanding literature in this field, though, it can be recognized that the published ANN-based studies only address either the FET DC current-voltage characteristics, also including dynamic trapping effects [14], [15], or the S-parameters [16], [17]. Large-signal FET models based on DNNs are instead at the pioneering level. Although various works focus on extracting the values of the equivalent large-signal circuit, little work has been dedicated on DNNs directly simulating the large-signal port waves.

This paper presents a pioneering methodology to model the nonlinear behavior of a FinFET transistor through incident and scattered waves. The proposed approach generalizes the X-parameter approach, that is in turn an extended version of S-parameters. Instead of extracting X-parameters from characterization, we exploit accurate TCAD simulations carried out by an in-house developed software [18], enabling the nonlinear device physical analysis through a drift-diffusion code implementing the Harmonic Balance algorithm. Here, TCAD-based X-parameters are used to train a DNN for nonlinear FinFET modeling [19], [20].

The advantage of the concurrent TCAD and ANN modelling is double: 1) a direct link to technological and physical parameters (e.g. FinFET geometry, temperature, etc.) can be seamlessly included in the ANN model to incorporate technological variability and temperature dependency; 2) compared to conventional X-parameters, which are ultimately a look-up table based model, ANNs large signal models don't

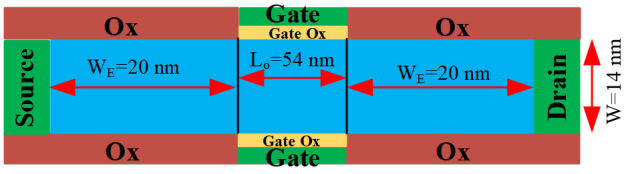


Fig. 1. Cross-section of the FinFET (individual finger) used for the X-parameter extraction [21].

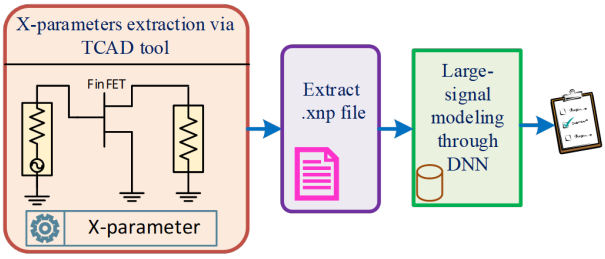


Fig. 2. General view of proposed methodology.

exploit data interpolation, hence DNNs require less memory allocation, lower numerical effort and allow faster and more accurate simulations.

The paper is organized as follows: Sec. II introduces the proposed approach. Sec. III reviews X-parameters and explains how the DNN is trained. Sec. IV explains the practical implementation of proposed LSTM- based DNN. Finally, Sec. V concludes this work.

II. Proposed Methodology In a Nutshell

The aim of this work is to extract a deep neural network (DNN) to mimic an FET device in large-signal operation.

To fix ideas, let us consider a FinFET device, used as the building block of a class A power amplifier operating at the frequency of 70 GHz for small-cells applications. The amplifier is designed assuming a multifinger device (10 fingers of 30 fins each) with a fin height of 25 nm, corresponding to a total gate periphery of 15 μm (only the two lateral channels for each fin are considered). The optimum DC bias has been selected at $V_{\mathrm{G}} = 0.675$ V and $V_{\mathrm{D}} = 0.6$V. The output port is terminated on the device optimum load (see [21] for details). The cross-section of each finger, reported in Fig. 1, has been simulated with our in-house TCAD simulator, at increasing level of input power, driving the device into nonlinear operation. At each power level, SS-LS analysis allows for the extraction of X-parameters [11], [21]. X-parameters could be directly plugged into circuit simulators (e.g. Keysight PathWave/ADS) for use in circuit analysis [22]. Here, instead, TCAD-based X-parameters are used to train a DNN for superior nonlinear FinFET modeling.

Fig. 2 presents the general extraction flow for our proposed method. X-parameters are extracted through the TCAD simulation at increasing level of input power and are then stored in a *.xnp* file. Then, a DNN is trained on the TCAD-based X-parameters, using the long short-term memory (LSTM), leading to predict the scattered waves for any incident wave. As the result, the extracted DNN effectively represents an equivalent model of the FinFET for large signal analysis, see Fig. 3, fully retaining the accuracy of the underlying physical analysis. Any variation of the technology can be further included in the DNN model. For example, in this preliminary work X-parameters are considered only at the temperature of 300 K, but the extension to a temperature-dependent DNN model, despite beyond the aim of this work, can be addressed simply retraining the DNN and adding an extra input neuron.

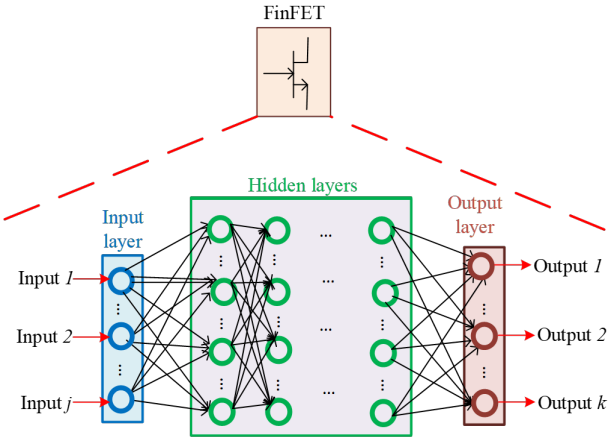


Fig. 3. DNN-based modeling of FinFET device. Input and output neurons represent the harmonic amplitudes of the incident and reflected waves, as in

III. DNN Training through X-parameters

X-parameters represent a particular approach to transistor large-signal modelling. It is generally understood that X-parameters are a generalization of S-parameters, since the input and output waveforms are divided into a *large-signal* part, corresponding to the transistor operating point, and a small-signal one, corresponding to the *wave variations*. Due to linearity, small perturbations are linearly linked. In the S-parameter case, the operating condition is static (DC) and the perturbations are the incident and reflected waves at frequency f adding to the DC point. In the X-parameter case, the working point itself is periodic, characterized by the fundamental frequency f_0 and harmonics nf_0. The perturbations are incident and reflected waves adding to the large-signal operating conditions at the same harmonics. Due to the nonlinearity, frequency conversion can occur among the harmonics and the device ports. Figure 4 shows the idea of X-parameters in terms of frequency components.

Matematically, (1) describes the development of the incident and reflected waves in terms of X-parameters.

$$B_{pm} = X^{(F)}_{pm}(|A_{11}|)P^m + X^{(S)}_{pm,qn}(|A_{11}|)P^{m-n}A_{qn} + X^{(T)}_{pm,qn}(|A_{11}|)P^{m+n}A^*_{qn} \quad (1)$$

where,

$$P = \frac{\mathrm{A}_{11}}{|A_{11}|} \quad (2)$$

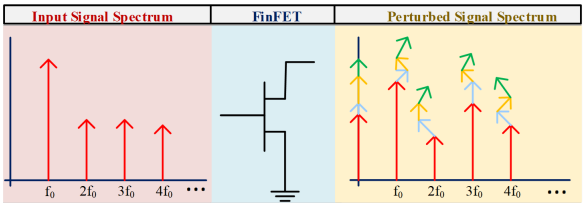


Fig. 4. The periodic large signal operating condition of a nonlinear device (left) is characterized by large amplitudes harmonics. The perturbation of the large signal operating point adds small contributions at each harmonic. X-parameters describe the link of the incident and reflected perturbations at each device port and frequency.

The reflected waves B_{pm} (labeled as port p and harmonic m) depend on the large signal input tone $|A_{11}|$ at port 1 and fundamental frequency, and on the small excitation tones A_{qn} (labeled as port q and harmonic n).

X-parameters include three sets of terms: X^F, X^S, and X^T. X^F describes the large signal harmonic operating condition; and X^S with X^T capture the small signal variations around the large-signal working point, linking the incident and scattered waves [23].

Turning to the DNN model extraction, we select the input layer which includes the incident waves (real and imaginary parts) A_{11}, A_{12}, ..., A_{15} for the first port and up to the fifth harmonic. Correspondingly, the output layer predicts the real and imaginary parts of scattered waves as B_{11}, B_{12} ,..., B_{15}, see Fig. 5. Notice that all the waves are divided into real and imaginary sections.

As the fundamental and important preliminary step for the DNN identification, a suitable amount of data is required for start training. Starting from the TCAD-based X-parameter data, a set of random incident waves A_{11}, A_{12}, ..., A_{15} is used in (1), obtaining the corresponding reflected waves B_{11}, B_{12} ,..., B_{15} (Step-3, Step-4). This dataset is divided into three parts as: training ($\mathrm{X_{Train}}$), validation ($\mathrm{X_{Val}}$), testing data ($\mathrm{X_{Test}}$) with a ratio of 70%, 15%, and 15%, respectively. For each of these input data, the corresponding output data can be presented as: $\mathrm{Y_{Train}}$, $\mathrm{Y_{Val}}$, and $\mathrm{Y_{Test}}$ as well.

Wwith these data the DNN is trained using the Eq. (3) (Step-5). Afterwards, the accuracy of DNN is calculated using (4) in MATLAB tool (Step-6).

$$\mathrm{net} = \mathrm{trainNetwork}(X_{\mathrm{Train}}, Y_{\mathrm{Train}}, \mathrm{layers}, \mathrm{options}) \quad (3)$$

$$Y_{\mathrm{Pred}} = \mathrm{predict}(\mathrm{net}, X_{\mathrm{Test}}) \quad (4)$$

The trained DNN is the regression one with LSTM hidden layers. The activation function is the rectified linear unit (ReLU) function where the normalized root mean square error (RMSE) is used for presenting the convergence of the proposed DNN.

For achieving the optimal hyperparameters, i.e., number of neurons and hidden layers, the rule of thumb is used (Step-7). By accurate training of DNN, the scattered wave can be predicted by any given incident wave (Step-8).

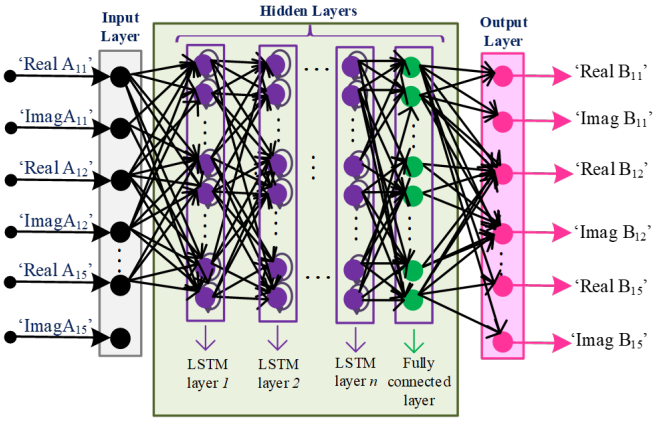


Fig. 5. Proposed LSTM-based DNN for large-signal modeling of FinFET transistors through X-parameters.

Algorithm 1 summarizes the employed steps for training and construing the regression LSTM-based DNN leading to model the FinFET transistor.

Algorithm 1 Summary of proposed large-signal modeling through DNN

1: X-parameter simulation environment in TCAD tool;
2: Extracting output file namely as '.xnp' file;
3: Preparing A_{11}, A_{12},...,A_{15} data series to be used in (1);
4: Calculating B_{11}, B_{12},...,B_{15} from (1);
5: Constructing the DNN structure includes input layer, hidden layers, and output layer as shown in Fig. 5;
6: Training the LSTM-based DNN by Eq. (3);
7: Optimizing the hyperparameters of DNN through the 'rule of thumb';
8: Validating the DNN output data through scattered waves for any given incident waves data.

IV. Practical Implementation of DNN

The proposed method is validated by implementing it in the CPU execution environment (Intel Core i7-4790 CPU @ 3.60 GHz and 32.0 GB RAM). As the first step, FinFET transistor namely as is inserted into the ADS environment.

As the initial step, the .xnp file including X^F, X^S, and X^T data for each port and harmonic is read by MATLAB. Afterwards, random A_{11}, A_{12},...,A_{15} are provided for calculating the output responses for the first port and harmonic B_{11}, B_{12},...,B_{15} from (1). In total 1000 random As includes real and imaginary parts are provided and respectively 1000 Bs are obtained. By using these data, the LSTM-based DNN is trained.

For updating the weights and biases of the DNN, adam optimization algorithm and standard gradient descent algorithm are used properly. The training options are set as solver to 'adam' and 'gradient threshold' to 1. The rule of thumb is employed for achieving the optimal hyperparameters. Fig. 6 shows the accuracy of trained DNN where in the third hidden layer at the 200^{th} neuron, the

Table 1. COMPARISON OF SELECTED SAMPLES OF WAVE HARMONICS RESULTING FROM THE XPAR MODEL VIA EQ.(1) AND FROM THE TRAINED DNN

	Xpar model via eq. (1)				**DNN model**			
Num of points	**Real** $B_{11}(\frac{v}{\sqrt{\Omega}})$	**Imag** $B_{11}(\frac{v}{\sqrt{\Omega}})$	**Real** $B_{12}(\frac{v}{\sqrt{\Omega}})$	**Imag** $B_{12}(\frac{v}{\sqrt{\Omega}})$	**Real** $B_{11}(\frac{v}{\sqrt{\Omega}})$	**Imag** $B_{11}(\frac{v}{\sqrt{\Omega}})$	**Real** $B_{12}(\frac{v}{\sqrt{\Omega}})$	**Imag** $B_{12}(\frac{v}{\sqrt{\Omega}})$
1	0.0067	0.0038	0.0100	0.0137	0.0073	0.0030	0.0122	0.0129
2	0.0090	0.0063	0.0060	0.0079	0.0083	0.0054	0.0065	0.0070
3	0.0147	0.0100	0.0057	0.0042	0.0203	0.0136	0.0069	0.0034
4	0.0229	0.0242	0.0265	0.0345	0.0314	0.0298	0.0301	0.0301
5	0.0179	0.0240	0.0379	0.0482	0.0136	0.0319	0.0291	0.0542
6	0.0275	0.0358	0.0522	0.0607	0.0203	0.0434	0.0501	0.0541

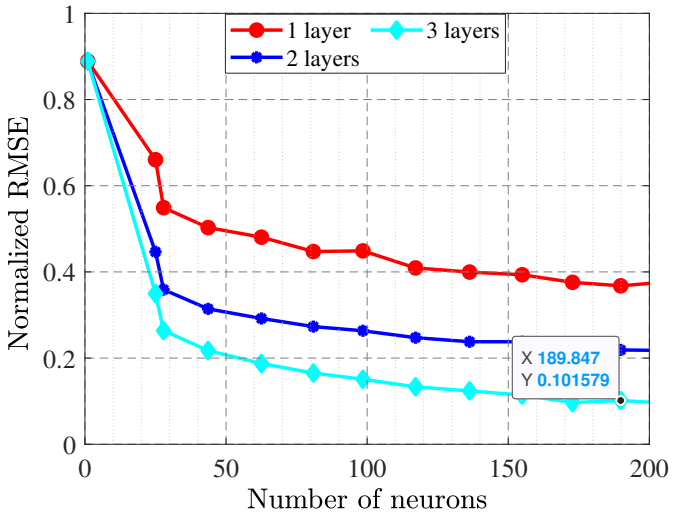


Fig. 6. Effects of hidden layer numbers in training an accurate DNN.

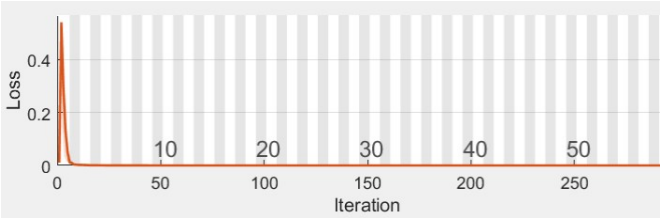


Fig. 7. Loss result of trained DNN over the iterations.

testing accuracy of proposed DNN in terms of normalized RMSE is around 0.10. Fig. 6 also shows that increasing the number of hidden layers the accuracy of DNN also increases. This suggests that in RF designs it is recommended to employ DNNs instead of shallow neural networks (SNNs). Additionally, the loss specification of trained DNN for 300 iteration is depicted in Fig. 7. Once the DNN is constructed and trained appropriately for any given incident waves (includes real and imaginary parts), the scattered waves in terms of real and imaginary sections can be predicted. Tab. 1 reports the comparison between the wave harmonics resulting from mathematical calculation through Eq. (1) and the trained DNN for a set of sample incident waves. The accuracy is quite good and proves that the proposed approach opens the way to develop reliable DNNs replacng X-pars.

V. CONCLUSION

This work presents an entirely new methodology for predicting the harmonic components of the waveforms in an electron device operated in Large Signal nonlinear conditions. In particular, akin to X-pars, the waveforms represent the variations with respect to a nominal operating condition, resulting either from a load mismatch or from a deviation of the device characteristics from a nominal device (nominal technology and material properties). The proposed approach exploits neural networs to replicate the Xpar device large signal model: this allows for a number of advantages in terms of numerical burden, speed and dependency on multiple physical parameters, such as temperature, device size and doping. As such, the proposed approach is a viable candidate to directly mimic the TCAD Large Signal simulations in terms of DNNs. In this preliminary paper, we have demonstrated how to provide a suitable amount of data to train the DNN: this is achieved by randomizing the incident waves and calculating the scattered waves through the X-parameter definition. The X-parameters (i.e., X^F, X^S, and X^T) have been obtained through accurate X-parameter simulation set-up in a TCAD environment. We trained a DNN by regression LSTM-based approach and demonstrated the accuracy of a DNN in comparison with a SNN, in terms of normalized RMSE and test samples. Once implemented into circuit simulators employing the Harmonic Balance algorithm, the proposed model will be a valuable nonlinear model for the microwave designer, including variability analysis and load sensitivity.

ACKNOWLEDGMENT

This work has been supported by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under the PRIN 2017 Project "Empowering GaN-on-SiC and GaN-on-Si technologies for the next challenging millimeter-wave applications (GANAPP)"

REFERENCES

[1] J. C. Bardin, D. H. Slichter, and D. J. Reilly, "Microwaves in quantum computing," *IEEE Journal of Microwaves*, vol. 1, no. 1, pp. 403–427, jan 2021.

[2] S. Callender, S. Pellerano, and C. Hull, "A 73ghz PA for 5G phased arrays in 14nm FinFET CMOS," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. IEEE, jun 2017.

[3] L. Kouhalvandi, "Directly matching an MMIC amplifier integrated with MIMO antenna through DNNs for future networks," *Sensors*, vol. 22, no. 18, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/18/7068

[4] R. Nikandish, "Gan integrated circuit power amplifiers: Developments and prospects," *IEEE Journal of Microwaves*, vol. 3, no. 1, pp. 441–452, 2023.

[5] D. Gruber, M. Clara, R. S. Pérez, Y.-S. Wang, C. Duller, G. Rauter, P. Torta, G. Knoblinger, and K. Azadet, "A 12-b 16-GS/s RF -sampling capacitive DAC for multi-band soft radio base-station applications with on-chip transmission-line matching network in 16-nm FinFET," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 12, pp. 3655–3667, 2021.

[6] J. Kim, S. Kundu, A. Balankutty, M. Beach, B. C. Kim, S. T. Kim, Y. Liu, S. K. Murthy, P. Wali, K. Yu, H. S. Kim, C.-C. Liu, D. Shin, A. Cohen, Y. Segal, Y. Fan, P. Li, and F. O'Mahony, "A 224-gb/s dac-based pam-4 quarter-rate transmitter with 8-tap ffe in 10-nm finfet," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 1, pp. 6–20, 2022.

[7] L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated Deep Neural learning-based optimization for high performance high power amplifier designs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 4420–4433, 2020.

[8] J. P. Duarte, S. Khandelwal, A. Medury, C. Hu, P. Kushwaha, H. Agarwal, A. Dasgupta, and Y. S. Chauhan, "BSIM-CMG: Standard FinFET compact model for advanced circuit design," in *ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC)*. IEEE, sep 2015.

[9] Verspecht and D. Root, "Polyharmonic distortion modeling," *IEEE Microwave Magazine*, vol. 7, no. 3, pp. 44–57, jun 2006.

[10] D. Root, J. Horn, J. Verspecht, and M. Marcu, *X-Parameters*. Cambridge University Press, 2013.

[11] S. Donati Guerrieri, F. Bonani, and G. Ghione, "Linking X parameters to physical simulations for design-oriented large-signal device variability modeling," in *2019 IEEE MTT-S International Microwave Symposium (IMS)*, 2019, pp. 204–207.

[12] C. Wilson, A. Zhu, J. Cai, and J. B. King, "Pade-approximation based behavioral modeling for RF power amplifier design," *IEEE Access*, vol. 9, pp. 18 904–18 914, 2021.

[13] H. Luo, X. Yan, J. Zhang, and Y. Guo, "A neural network-based hybrid physical model for gan hemts," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4816–4826, 2022.

[14] A. Jarndal, "GaN HEMT electrothermal modeling using feedback neural networks technique," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2019, pp. 1–4.

[15] ——, "Gray wolf optimization-based modeling technique applied to gan high mobility electron transistors," *IEEE Journal of the Electron Devices Society*, vol. 9, pp. 958–965, 2021.

[16] A. Jarndal, S. Husain, and M. Hashmi, "Genetic algorithm initialized artificial neural network based temperature dependent small-signal modeling technique for GaN high electron mobility transistors," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 31, no. 3, p. e22542, 2021.

[17] A. Jarndal, "Neural network electrothermal modeling approach for microwave active devices," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 29, no. 9, p. e21764, 2019.

[18] S. Donati Guerrieri, F. Bonani, F. Bertazzi, and G. Ghione, "A unified approach to the sensitivity and variability physics-based modeling of semiconductor devices operated in dynamic conditions—Part I: Large-signal sensitivity," *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 1195–1201, mar 2016.

[19] E. Catoggio, S. Donati Guerrieri, C. Ramella, and F. Bonani, "Thermal modeling of RF FinFET PAs through temperature-dependent X-parameters extracted from physics-based simulations," in *2022 International Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMiC)*. IEEE, apr 2022.

[20] A. M. Bughio, S. Donati Guerrieri, F. Bonani, and G. Ghione, "Physics-based modeling of FinFET RF variability," in *2016 11th European Microwave Integrated Circuits Conference (EuMIC)*. IEEE, oct 2016.

[21] E. Catoggio, S. Donati Guerrieri, and F. Bonani, "Efficient TCAD thermal analysis of semiconductor devices," *IEEE Transactions on Electron Devices*, vol. 68, no. 11, pp. 5462–5468, nov 2021.

[22] S. Donati Guerrieri, C. Ramella, E. Catoggio, and F. Bonani, "TCAD-based dynamic thermal X-parameters for PA self-heating analysis," in *2022 17th European Microwave Integrated Circuits Conference (EuMIC)*. IEEE, sep 2022.

[23] X-PARAMETER generator. http://keysight.com. Accessed: 2023-01-15.