A Hybrid GaN HEMT Model Merging Artificial Neural Networks and ASM-HEMT for Parameter Precision and Scalability

Zhongzhiguang Lu, *Member, IEEE*, Hanchao Li, Hanlin Xie, Yihao Zhuang, Wang Wensong, *Senior Member, IEEE*, Ng Geok Ing, and Yuanjin Zheng, *Senior Member, IEEE*

***Abstract*—An innovative hybrid physical model for Gallium Nitride High-Electron-Mobility Transistors (GaN HEMTs) that leverages an artificial neural network (ANN) approach is proposed. This model utilizes ANN to formulate surface potentials, correlating them with the device's RF and DC behaviors in accordance with the Advanced SPICE Model for GaN HEMTs (ASM-GaN-HEMT) theoretical framework. The extraction process is refined through Multi-Objective Particle Swarm Optimization (MOPSO), enhancing the precision of the extracted parameters. Subsequently, a two hidden-layer ANN architecture is employed to derive the surface potential at the source and drain terminals (ψ_s, ψ_d). These surface potentials form the basis of the hybrid model within the ASM-HEMT framework, including the trapping and self-heating effects. Validation of the model is conducted using the Advanced Design System (ADS) simulation platform. The hybrid ANN-based model exhibits scalability and accuracy in comparison to traditional ASM-based physical models. Experimental validations demonstrate a strong concordance between the hybrid model's predictions and the empirical data across a range of tests, including IV characteristics, S-parameters, and load-pull power sweeps. The proposed method significantly improves the accuracy of the S parameter compared to traditional models, reducing the large signal performance error to within 5%. The results show the robustness of the proposed model and its potential to enhance the predictive modeling capabilities for GaN HEMT devices.**

***Index Terms*—Gallium Nitride High-Electron-Mobility Transistors (GaN HEMTs), Advanced SPICE Model for GaN HEMTs (ASM-GaN-HEMT), device modeling, parameter extraction, neural network.**

Manuscript received 14 May 2024; revised XXXX; accepted XXXX. Date of publication XXXX; date of current version XXXX. This work is supported by the National Research Foundation and A*STAR under the RIE2025 Manufacturing, Trade & Connectivity (MTC) Industry Alignment Fund Pre-Positioning (IAF-PP) (Grant No: M22L3a0112)

Zhongzhiguang Lu, Hanchao Li, Yihao Zhuang, Wang Wensong, Geok Ing Ng and Yuanjin Zheng are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: luzh0013@e.ntu.edu.sg; yjzheng@ntu.edu.sg).

Hanlin Xie and Geok Ing Ng are with National GaN Technology Centre (NGTC), and Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), Innovis #08-02, Singapore 138634, Republic of Singapore.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

I. Introduction

Gallium nitride (GaN) HEMT stands at the forefront of semiconductor technology, particularly in microwave circuit design. Their advantages are manifold, including unparalleled power density, operation capabilities extending to millimeter-wave frequencies, and exceptional efficiency—attributes owed to their high electron mobility, elevated cutoff frequency, and minimized on-resistance and capacitance [1], [2]. GaN device modeling is crucial in designing circuits as it accurately predicts device performance, allows for circuit simulation and optimization, and enables parameter extraction essential for accurate simulation and optimization [3]-[5]. The complexity of the electrical behavior of GaN HEMT devices poses a challenge in accurately predicting their performance, necessitating a robust device model.

GaN device modeling methodologies are diverse, spanning empirical [6], [7], physics-based [8]-[10], and artificial neural network (ANN) models [11]-[13]. Empirical models rely on measured data and are simple to implement but lack physical insight. Physics-based models are more accurate but computationally expensive [14]-[16]. ANN based GaN device modeling is also a popular approach that uses machine learning techniques to create accurate models without the need for extensive physical knowledge. ANN models can capture non-linear and complex device behaviors like self-heating and trapping effects and are computationally efficient [17], [18]. Nonetheless, their effectiveness hinges on the availability of large, high-quality datasets for training, and their accuracy is occasionally constrained by the caliber of input data. Recent years have seen a surge in the application of ANN in GaN HEMT modeling. Research has explored various innovative approaches, such as evolutionary multilayer perceptron (EMLP) based modeling [19], gate charge model based on neural networks [20], and electrothermal modeling approach using ANN applied to GaN devices [21]. Further research has explored machine learning techniques for temperature-dependent small signal modeling [22]-[24], the comparative analysis of varied ANN architectures for behavioral modeling [25], and the innovative use of a hybrid multilayer perceptron-random forest model for meticulous estimation of device parameters [26]. Additionally, novel methodologies like Wiener-type dynamic neural network (DNN) structures for nonlinear device modeling have also been put forth [27]. Although ANN models adeptly replicate device behavior, their lack of physical interpretability obscures the understanding of underlying physical principles. This gap has catalyzed the exploration of hybrid models combining ANN and physical modeling approaches [28]. Such hybrid models promise enhanced accuracy and computational efficiency over standalone ANN models. They adeptly capture complex physical phenomena while curtailing the volume of training data required.

This article introduces an innovative hybrid physical model that leverages ANN to precisely characterize the source-side and drain-side surface potentials (ψ_s, ψ_d) of the GaN HEMT, which are the key parameters in the ASM-GaN-HEMT model [8], [10], [29]-[33]. In addition, this article employs parameter extraction and multi-objective particle swarm optimization (MOPSO) methodology designed for the hybrid model [34]. Moreover, this hybrid model exhibits enhanced scalability relative to traditional ANN-based models [35], as evidenced by its integration of dimensional parameters within the modeling process. This aspect significantly reduces the modeling cost for devices sharing the same manufacturing process but differing in physical characteristics. The methodology presented is especially advantageous for the precise modeling of a collection of devices showcasing a variety of physical attributes, including size variations.

This article is organized as follows. In Section II, the parameter extraction and optimization strategies for this hybrid model are elaborated. In Section III, the ANN-based hybrid physical model is built, whose core model includes the ANN-based surface potential model and the ASM-based model equations. Section IV uses multiple ways to verify the performance of the proposed hybrid model, including the current–voltage (IV), S-parameters, and large-signal tests. The conclusion is made in Section V.

II. Parameter Extraction for Model Implementation

In this section, we detail the structural and electrical characterization of AlN/GaN-on-Si High Electron Mobility Transistors (HEMTs), shown in Fig. 1(a), (c) and (d), which combines a cross-sectional schematic and a Transmission Electron Microscopy (TEM) image of the fabricated device. Fabricated on a 6" high-resistivity (>6000 Ω.cm) Si (111) substrate, the III-nitride materials were epitaxially grown via Metal-Organic Chemical Vapor Deposition (MOCVD). The epitaxial stack includes a 1.5 μm GaN buffer, a 6 nm AlN barrier, and a 4 nm in-situ SiN cap layer, which together enable a 2-dimensional electron gas (2DEG) with a density of 2.06 × 10^{13} cm^{-2}, electron mobility of 1060 cm^2/Vs, and a sheet resistance of 285 Ω/□, as derived from Hall measurements.

The device fabrication commenced with mesa isolation using Cl2/BCl3 plasma dry etching to achieve a mesa height of approximately 120 nm. Subsequent steps involved the formation of ohmic contacts by the deposition of a Ti/Al/Ni/Au metal stack with respective thicknesses of 20/120/40/50 nm, of 20/120/40/50 nm, and rapid thermal annealing at 775°C for 30 seconds in an N_2 atmosphere. The definition of T-gates, with a 190 nm gate foot, was performed using electron beam lithography with a bilayer e-beam resist of PMMA/MMA, followed by the deposition of Ni/Au (30/300 nm) and a lift-off process. To mitigate current collapse and enhance device

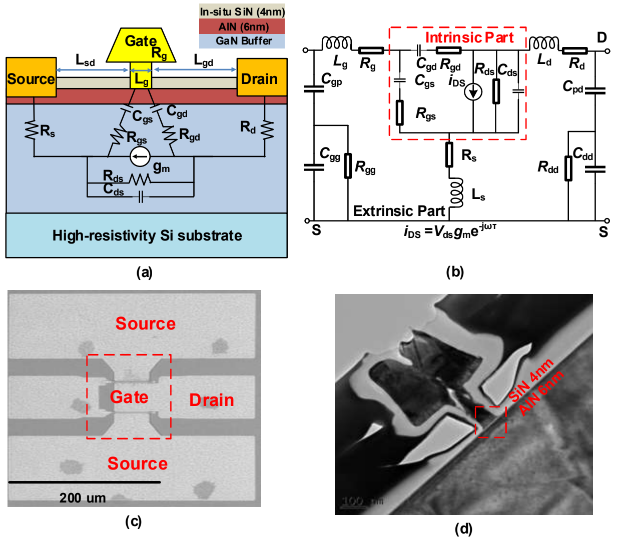


Fig. 1. (a) Cross-sectional diagram of the AlN/GaN-on-Si HEMTs; (b) Equivalent circuit for GaN HEMT, (c) TEM images of top view, (d) TEM images of gate and interface.

reliability, the surface was passivated with a 20 nm SiN layer deposited by PECVD at 300°C, complemented by an additional ex situ PECVD SiN layer due to the critical thickness limitations of the in-situ SiN.

Based on the in-house fabricated GaN HEMT, this article innovatively synthesizes the ASM-GaN-HEMT Model with ANN technology, proposing a hybrid physical and ANN-based GaN HEMT model. The ASM-GaN-HEMT model, a compact and surface-potential-based framework, is designed for GaN HEMTs and adeptly evaluates the quasi-Fermi level as a function of bias, having met the rigorous benchmarks of the CMC standard. A key aspect of this model is the precise calculation of intrinsic capacitances. In this framework, the intrinsic gate (Q_{gi}) and drain (Q_{di}) charges are determined based on the surface potential at both ends of the channel (ψ_s, ψ_d).

The hybrid physical and ANN-based GaN model method proposed here is applied. In ASM-GaN-HEMT model, due to the difference in barrier thickness under the filed plate, the Voff will be different from the intrinsic transistor, Since the GaN-on-Si HEMTs to be modeled in this paper are designed for low voltage 5G applications, there is no filed plate in the structure, which will simplify the three equivalent transistors in ASM into one equivalent transistor in this paper.

Fig. 1(b) depicts the equivalent small-signal circuit diagram of the in-house GaN HEMT. To address the unique RF characteristics of the Si substrate, especially under low-frequency pinch-off conditions, an RC branch comprising parameters such as C_{gg}, C_{dd}, R_{gg}, R_{dd}, C_{gp}, C_{dp} is integrated [18], alongside other critical parameters including connection inductances (L_s, L_g, L_d), access region resistances (R_s, R_g, R_d), intrinsic capacitances (C_{gs}, C_{gd}, C_{ds}), transconductance (g_m), and output conductance (G_{ds}). Fig. 2 summarizes the process flow for extracting and optimizing surface potentials (ψ_s, ψ_d) in the hybrid model, the average computational time of each process is listed in the right. The journey commences with extracting pad parameters from the S-parameter measurements under pinch-off bias conditions (C_{gg}, C_{dd}, R_{gg}, R_{dd}, C_{gp}, C_{dp}). Subsequently, the parasitic parameters (L_s, L_g, L_d, R_s, R_g, R_d) are derived from the S-parameter results under unbiased conditions (V_{GS} = V_{DS} = 0V). After de-embedding the extrinsic components of the S parameters under different bias condition, the values of the intrinsic components can be calculated by the equations in [36]-[38].

It is imperative to acknowledge that the intrinsic capacitances within the GaN HEMT model, specifically C_{gs}, C_{gd}, C_{ds}, exhibit variations in response to changes in the extrinsic voltages V_{GS} and V_{DS}. To accurately derive the intrinsic charge model and the nonlinear drain-source current model, a crucial step involves the elimination of the voltage drops across the parasitic resistances, namely R_g, R_d and R_s from the extrinsic voltages V_{GS} and V_{DS}. This adjustment is essential to accurately determine the intrinsic voltages, V_{gs} and V_{ds}. This expression listed below plays a pivotal role in ensuring that the model accurately reflects the intrinsic electrical characteristics of the GaN HEMT.

$$\begin{bmatrix} V_{gs} \\ V_{ds} \end{bmatrix} = \begin{bmatrix} V_{GS} \\ V_{DS} \end{bmatrix} - \begin{bmatrix} R_g + R_s & R_s \\ R_s & R_d + R_s \end{bmatrix} \begin{bmatrix} I_{gs} \\ I_{ds} \end{bmatrix} \tag{1}$$

Numerical integration of the intrinsic capacitance C_{gs} and C_{gd} as a function of the intrinsic voltage enables the derivation of charge expression. In alignment with the ASM-GaN-HEMT model description and reflecting the specific characteristics of our in-house GaN HEMT, the expressions for intrinsic gate charge (Q_{gi}) and drain (Q_{di}) charges are as follows [8].

$$Q_{gi} = WLC_g \left\{ V_{go} - \frac{1}{2}(\psi_s + \psi_d) + \frac{1}{12}\frac{(\psi_d - \psi_s)^2}{V_{goT}} \right\}$$

$$Q_{di} = -\frac{1}{2} WLC_g \left\{ \begin{array}{c} V_{go} - \frac{1}{3}(\psi_s + 2\psi_d) + \\ \frac{1}{12}\frac{V_{DS}^2}{V_{goT}}\left(1 + \frac{1}{10}\frac{V_{DS}}{V_{goT}}\right) + \frac{V_{DS}^3}{120 V_{goT}^2} \end{array} \right\} \tag{2}$$

where W and L means the gate width and length of the GaN HEMT, C_g denotes the gate capacitance per unit area, $C_g = \frac{\varepsilon}{d}$. V_{go} symbolizes the gate override voltage V_{GS}−V_{off}, and the expression of V_{goT} is given below, and K_B is the Boltzmann constant.

$$V_{goT} = V_{go} + \frac{K_B T}{q} - \frac{1}{2}(\psi_s + \psi_d) \tag{3}$$

MATLAB is employed to compute the source-side and drain-side surface potentials ψ_s and ψ_d using the equations (2). Yet, these initially calculated values require fine-tuning to negate influences from non-ideal factors such as instrumental errors or interference. In adherence to the ASM-GaN-HEMT model's theoretical framework, the conditions that $\psi_s = E_F$ and $\psi_d = E_F + V_{ds}$, must be satisfied, ensuring ψ_d exceeds ψ_s whenever V_{ds} is positive. The refined values of ψ_s and ψ_d should comply with these rules, signifying their critical role within the ASM-GaN-HEMT model by impacting several device characteristics. The surface potential dictates the drain current by modulating the transconductance g_m and influences the high-frequency response through its effects on intrinsic capacitances. This article seeks to establish precise models for both DCIV and RF performance of GaN HEMTs, utilizing ANNs. Therefore, as shown in Fig. 2, the value of ψ_s is derived from the intrinsic gate and drain charges. ψ_d is then

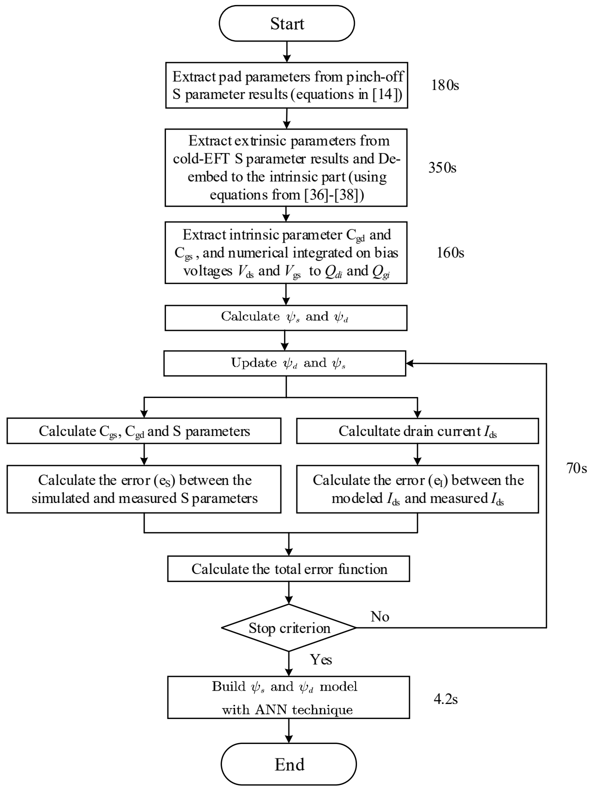


Fig. 2. Optimization flowchart of ψ_s and ψ_d.

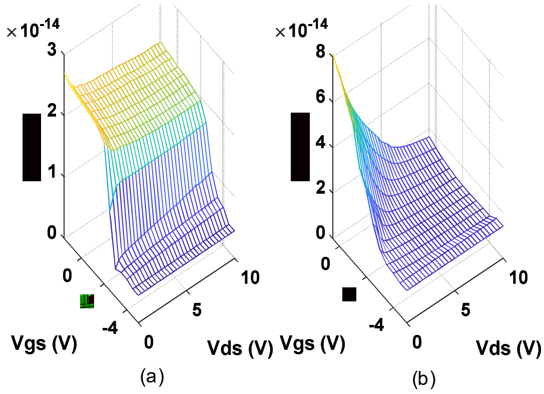


Fig. 3. The derived (a) C_{gs} and (b) C_{gd} under different bias conditions.

computed by incrementing ψ_s by the drain-to-source voltage (V_{DS}), aligning the model with the applied bias conditions. Subsequently, two parallel streams of calculations proceed. The first stream involves the computation of the gate-to-source (C_{gs}) and gate-to-drain (C_{gd}) capacitances, key parameters that affect the transistor's frequency response. Following this, the S-parameters are obtained, which are essential for characterizing the high-frequency behavior of the device. An error metric (e_S) is then calculated by quantifying the deviation between the simulated S-parameters and the empirical S-parameters obtained from measurements.

Concurrently, the second stream addresses the direct current characteristics by obtaining the drain current (I_{ds}) from the model. The value of drain current I_{ds} could be calculated using the physical equations (7) given in Section III. The error (e_I) in this stream is determined by the difference between the modeled I_{ds} and the measured I_{ds}, which reflects the model's fidelity in replicating the device's DC performance. e_I and e_S from the parallel streams are then amalgamated to formulate a total error function, $e_{total} = w_I e_I + w_S e_S$. w_I and w_S are the weights of e_I and e_S, both are 0.5 in this paper. This composite error function serves as an indicator of the overall discrepancy between the model's predictions and the measured data across both frequency-dependent and static operational regimes.

The iterative cycle involves updating ψ_s and ψ_d to minimize the total error function, refining the model with each iteration. This process is repeated until a predefined stopping criterion is satisfied. Such a criterion could be a minimal threshold for the error function or a maximum number of iterations, ensuring convergence to an optimal solution within practical computational limits.

Based on the principles of the particle swarm optimization (PSO) algorithm, the MOPSO [34] is a nature-inspired metaheuristic algorithm used to solve multi-objective optimization problems, which has many advantages, such as diversity preservation, efficient exploration. Like PSO, particles in MOPSO adjust their positions based on their own best position found so far and the global best position found by the entire swarm. These updates aim to guide the search toward better solutions in the multi-objective space. In multi-objective optimization, the goal is to find a set of solutions that represent a trade-off between different objectives, forming a Pareto front. In this work, MOPSO is employed to optimize the value of ψ_s and ψ_d for global accuracy. The optimization is carried out in Matlab. Then, the ANN-based ψ_s and ψ_d model is illustrated in Section III.

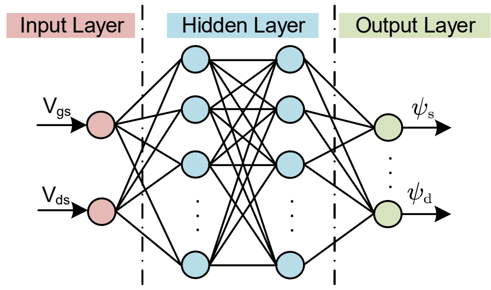


Fig. 4. Two hidden layers ANN framework for ψ_s and ψ_d model.

TABLE I COMPARISON OF TRAINING TIME AND MSE FOR DIFFERENT NEURAL NETWORK CONFIGURATIONS

Neuron numbers	Computational Time	MSE
7,5	3.58s	0.0551
9,7	4.23s	0.0275
11,9	5.4s	0.0159

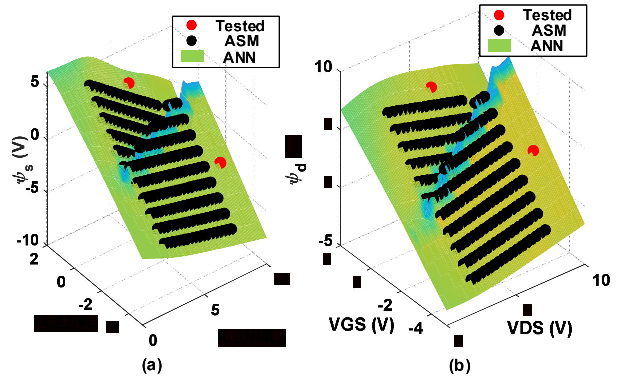


Fig. 5. ASM-based (black dots) and ANN-based (surface) results of (a) ψ_s and (b) ψ_d. V_{DS} ranges from 0 to 10 V and V_{GS} from −5.0 to 1.5 V.

III. GaN Device Modeling Process

A. Surface Potential Modeling

In this section, a neural network architecture featuring two hidden layers is utilized to encapsulate the nonlinear interdependencies between the surface potential ψ_s, ψ_d and the intrinsic voltages V_{ds}, V_{gs}. This approach diverges from the conventional lookup table-based large-signal models, where intrinsic voltages are discretized at uniformly distributed intervals. Due to the non-uniform spacing of the intrinsic voltages in this context, interpolation methods are necessitated to facilitate the rasterization of V_{gs} and V_{ds} In contrast, when leveraging ANNs for the modeling of large-signal behaviors, it becomes feasible to train the model directly using the intrinsic voltages V_{gs} and V_{ds}, thereby obviating the need for rasterization. The ANNs inherently select either a sigmoid or a linear activation function, which ensures the preservation of data continuity and the retention of higher-order differentiability within the model. Consequently, the formulations for the ANN-based ψ_s and ψ_d are posited as follows, integrating the considerations to provide a robust framework for the predictive simulation of transistor characteristics.

The tanh function is selected due to its effectiveness in modeling non-linearities and its smooth, bounded output, which helps stabilize the training process. The tanh function is known to provide better performance compared to other activation functions like the sigmoid, especially when dealing with both positive and negative values. This property makes tanh particularly useful for this application, as it allows for more symmetric output scaling, which was necessary for accurately capturing the variations in our dataset. Moreover, the tanh function has been widely adopted in similar non-linear modeling tasks because of its ability to approximate complex behaviors while maintaining smoothness and differentiability, making it an ideal candidate for the proposed neural network model [20], [21], [39].

It is well-established that neural networks with one or more hidden layers are effective at capturing complex, non-linear relationships between input and output variables. We opted for two hidden layers in this case because, for the specific complexity of the dataset and the target application, this architecture provides an adequate balance between model accuracy and computational efficiency. While a single hidden layer can model non-linear relationships, multiple hidden layers enable the network to learn hierarchical features more effectively, which improves its ability to model complex functions. This configuration utilizes a backpropagation (BP) neural network with dual hidden layers [39-41]. Therefore, the functional representation of ψ_s and ψ_d as a response to V_{gs} and V_{ds} could be articulated as follows.

$$\psi_k(V_{gs}, V_{ds}) = b_3 + \sum_{i=1}^{M} w_{3i} \tanh(\varphi_{2i}) \tag{4}$$

Where φ_{2i} is defined by

$$\varphi_{2i} = b_{2i} + \sum_{j=1}^{N} w_{2ij} \tanh(\varphi_{1j}) \tag{5}$$

Where φ_{1j} is specified by

$$\varphi_{1j} = b_{1j} + w_{1j1} V_{\mathrm{gs}} + w_{1j2} V_{\mathrm{ds}} \tag{6}$$

Here, φ_{2i} and φ_{1j} are the output function of the second and the first hidden layer, respectively. Specifically, we trained and evaluated three different configurations of the neural network: (7, 5), (9, 7), and (11, 9). The training time for each configuration and the corresponding Mean Squared Error (MSE) are provided in the Table I. The (7, 5) configuration required 3.6 seconds, (9, 7) took 4.2 seconds, and (11, 9) took 5.4 seconds. As demonstrated, increasing the number of neurons improves the fitting performance but also increases the computational time.

However, selecting too many neurons can significantly increase the complexity of the system, extend the simulation time on the ADS platform, and lead to overfitting. After considering the trade-off between model complexity and fitting performance, we determined that the (9, 7) configuration offers the best balance between these factors. This selection allows for improved accuracy while maintaining reasonable computational efficiency. The notations w and b signify the weights and biases of the neurons within the hidden layers and the output layer, respectively. A total of 714 observations are employed in the neural network, of which 70% were allocated to the training set 15% to the validation set and 15% to the testing set. This split ensured that the model was trained on enough data while also maintaining a robust validation process to evaluate its performance. The data was initially cleaned by removing any outliers or missing values to ensure data integrity. Afterward, normalization was applied to rescale the input features to a range between [0, 1], which is a common practice to improve neural network performance. The convergence criterion of the neural network is based on the minimization of the MSE on the validation set. The training process was set to run for a maximum of 10000 epochs, but early stopping was employed to prevent overfitting. The metric was the MSE for both the training and validation sets, providing a measure of the model's accuracy [39-41].

The technique employed here constructs a nonlinear connection between the surface potential and drain potential, noting that the trained ANN model does not convey any direct physical meaning but rather serves as a computational tool for simulation and prediction. To make the trained ANN model conform to some physical rules of the surface potential model, such as the ψ_{s} and ψ_{d} will increase with the increasing V_{gs}, the ANN models need to be trained multiple times and carefully selected.

Fig. 5 (a) and (c) show the ASM-based and ANN-based ψ_{s} and ψ_{d} results in different bias condition. V_{gs} is selected from -5 to 1.5V with 0.25V step and V_{ds} is selected from 0 to 10V with 0.5V steps.

B. Drain Current Modeling

Within the framework of the ASM-GaN HEMT model, the drain current I_{ds} is expressed by the following equation.

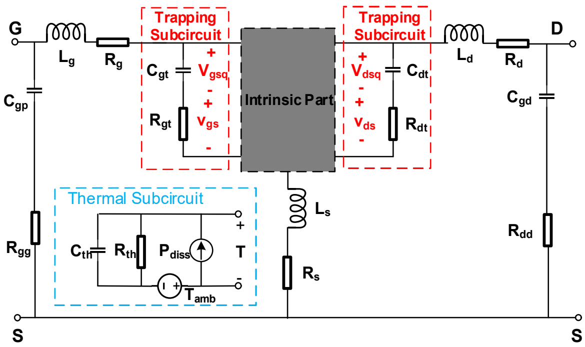


Fig. 6. Large signal equivalent circuit for GaN HEMTs with thermal and trapping subcircuits.

$$I_{\mathrm{ds}} = \frac{\mu_{\mathrm{eff}}}{\sqrt{1+\theta_{\mathrm{sat}}^2 V_{\mathrm{ds}}^2}} \frac{W}{L} C_{\mathrm{g}} N_{\mathrm{f}} V_{\mathrm{goT}} V_{\mathrm{ds}} (1 + \lambda V_{\mathrm{ds}}) \tag{7}$$

Where μ_{eff} symbolizes the effective mobility, and θ_{sat} is a fitting parameter. N_{f} represents the number of fingers, and λ signifies the channel length modulation coefficient.

The effective mobility μ_{eff} is further described by the relation.

$$\mu_{\mathrm{eff}} = \frac{\mu_0}{1 + UA \cdot E_{\mathrm{avg}} + UB \cdot E_{\mathrm{avg}}^2} \tag{8}$$

where E_{avg} is the average vertical electric field computed as

$$E_{\mathrm{avg}} = \frac{V_{\mathrm{gs}} - \psi_{\mathrm{m}}}{T_{\mathrm{bar}}} \tag{9}$$

In this context, UA and UB are parameters from the ASM-HEMT model, with typical values being in the order of 10^{-9} and 10^{-18}, respectively. E_{avg} denotes the average vertical electric field and ψ_{m} is the average surface-potential which is obtained by averaging the surface potential at the source and the drain ends of the channel. T_{bar} is the barrier layer thickness.

C. Nonlinear effects Modeling

In GaN HEMTs, two critical nonlinear phenomena, self-heating and trapping effects, are observed. These effects are instrumental in influencing the operational stability and efficiency of the devices. The self-heating effect is caused by the power dissipation within the device, leading to a temperature rise that affects the electrical characteristics. The trapping effect, on the other hand, refers to the capture of charge carriers in the gate area, modifying the threshold voltage and, consequently, the drain current.

Self-heating refers to the thermal elevation within the transistor structure due to power dissipation, which can lead to a degradation of electron mobility and a shift in device characteristics. To capture this phenomenon, an RC network, comprising a thermal resistance (R_{th}) and a thermal capacitance (C_{th}) is employed. The equations are listed below.

$$\begin{gathered} i_{\mathrm{DS}} = I_{\mathrm{ds}} \times (1 + K_{\mathrm{T}} \Delta T) \\ K_{\mathrm{T}} = \alpha_0 + \alpha_1 [1 + \tanh(\alpha_2 + \alpha_3 V_{\mathrm{ds}})] \\ T = T_{\mathrm{ref}} + \Delta T,\ \Delta T = P_{\mathrm{diss}} R_{\mathrm{th}} + T_{\mathrm{amb}} - T_{\mathrm{ref}} \\ \tau_{\mathrm{th}} = R_{\mathrm{th}} C_{\mathrm{th}} \end{gathered} \tag{10}$$

The self-heating effect is represented by modifying the drain current I_{ds} with a factor (K_{T}), and α_0, α_1, α_2 and α_3 are fitting parameters. This configuration accounts for the temporal evolution of the device temperature due to power dissipation. The elevation in temperature at the thermal node is quantified by a temperature increase (ΔT), which is superimposed on a predefined reference temperature, $T_{ref} = 25$°C. The product of R_{th} and C_{th} defines the thermal time constant τ_{th}, which dictates the rate of thermal transients within the device. C_{th} is selected to establish a thermal time constant of approximately 0.1ms, ensuring that the model captures the rapid thermal dynamics characteristic of GaN HEMTs.

The trapping effect in GaN HEMTs, which involves charge carriers getting caught in defects, presents a significant design challenge by altering device performance, causing issues like threshold voltage shifts and current collapse. To measure this effect and obtain accurate device characteristics, pulsed I-V testing is employed. This method applies voltage pulses to minimize self-heating and avoid the impact of charge trapping, offering a clearer view of the transistor's behavior. This technique is essential for developing models that predict device behavior and ensures the reliability of GaN HEMTs in high-frequency, high-power applications.

The trapping effect are quantitatively accounted for by incorporating two resistive-capacitive (RC) subcircuits, shown in Fig. 6. These subcircuits are designed to emulate the alterations in the gate- source voltage induced by the trapping effects within both the surface states and the subsurface buffer layer [18].

$$\begin{aligned} V_{GS\_eff} &= V_{gs} + \gamma_{surf} + \gamma_{subs} \\ \gamma_{surf} &= k_{g1}\left(V_{gs} - V_{gsq0}\right)\left(V_{gsq} - V_{gsq0}\right) \\ \gamma_{subs} &= \begin{cases} \tanh\left(k_{g2}\left(V_{ds} - V_{dsq}\right)\left(V_{gs} - V_{gsq1}\right)\right) & \\ + k_{g3}V_{dsq}\left(V_{gs} - V_{gsq2}\right), & V_{ds} \leq V_{dsq} \\ k_{g3}V_{dsq}\left(V_{gs} - V_{gsq2}\right), & V_{ds} \leq V_{dsq} \end{cases} \end{aligned} \tag{11}$$

The effective gate-source voltage V_{GS_eff} is formulated by adjusting the gate-source voltage V_{gs} with two terms γ_{surf} and γ_{sub}, which represent the surface and buffer-related trapping effects, respectively. The fitting parameters k_{g1}, k_{g2} and k_{g3} are calibrated to encapsulate the complex dynamics of charge trapping and de-trapping, which are instrumental in the accurate prediction of the device's transient and steady-state responses. The quiescent voltages V_{gsq0}, V_{gsq1} and V_{gsq2} serve as the reference points for these trapping-induced potential shifts. Through this meticulous formulation, the model aspires to closely replicate the GaN HEMT's performance characteristics, accounting for the nuanced interplay between electrical operation and the trapping mechanisms intrinsic to the device architecture.

D.Gate Current Modeling

The gate currents I_{gs} and I_{gd} prensent the behavioral characteristics of the Schottky contacts present in the device architecture. These currents can be accurately modeled by either utilizing an artificial neural network that captures the nonlinearities associated with the Schottky barrier or by employing a traditional Schottky diode model equation. The latter is mathematically represented as below [42].

$$I_{gs} = I_{sgs}\left(e^{\frac{qV_{gd}}{kTN_{gs}}} - 1\right), \quad I_{gd} = I_{sgd}\left(e^{\frac{qV_{gd}}{kTN_{gd}}} - 1\right) \tag{12}$$

Where q is the charge of an electron, k is the Boltzmann constant, T is the absolute temperature, and N_{gs} and N_{gd} are the ideality factors of the gate-source and gate-drain Schottky diodes, respectively. I_{sgs} and I_{sgd} are the saturation currents for the gate-source and gate-drain junctions, respectively, and

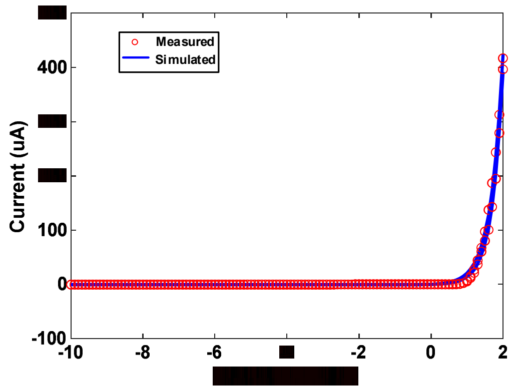


Fig. 7. The measured (red symbols) and simulated (blue line) gate current of the 2 × 50 μm device.

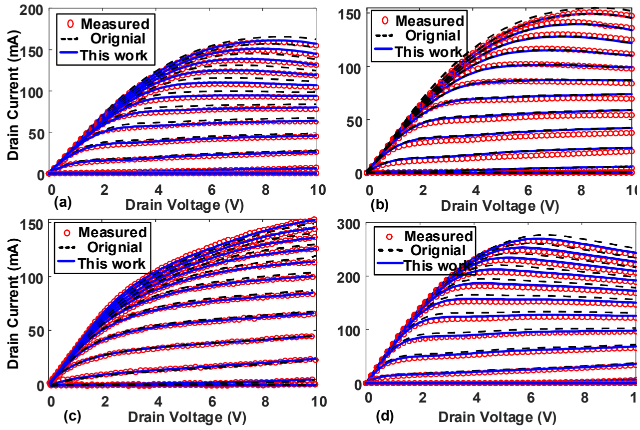


Fig. 8. Measured (red symbols), ASM model simulated (black dashed lines) and hybrid model simulated (blue lines) IV characteristics of (a) 2 × 50 μm device, DCIV at 25°C, (b) 2 × 50 μm device, DCIV at 65°C, (c) 2 × 50 μm device, PIV testing, V_{dsq} = -5 V, V_{gsq} = 10 V, (d) 2 × 100 μm device, DCIV at 25°C.

serve as fitting parameters to align the model with empirical observations. Fig.7 has shown the measured and simulated results of gate current of the 2 × 50 μm device. These expressions are instrumental in defining the gate leakage behavior under different biasing conditions and are crucial for the prediction of the off-state leakage currents in GaN HEMT devices.

However, at extremely high voltages, where non-linear field effects and potential breakdown mechanisms become significant, the assumptions made in the proposed model may require refinement. In particular, the absence of breakdown consideration means the model may not be suitable for voltage levels near or beyond the device's critical breakdown voltage.

IV. Model Verification

In this paper, a comprehensive suite of experimental evaluations—namely DCIV, Pulsed IV (PIV), and load-pull measurements—were conducted on two in-house fabricated GaN HEMT devices. The two GaN HEMT devices both use in-situ technology with 100-nm gate length and 100-μm total gate width (2 fingers × 50 μm), 200-μm total gate width (2 fingers × 100 μm), respectively. Employing the outcomes from these experimental evaluations and the proposed modeling

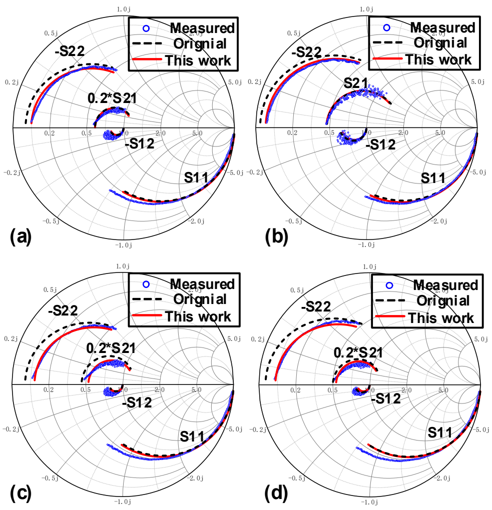


Fig. 9. Measured (blue symbols), ASM model simulated (black dashed lines) and hybrid model simulated (red lines) S-parameters in 1-40 GHz of (a) 2 × 50 μm device, $V_{GS} = -2.0$ V, $V_{DS} = 10$ V; (b) 2 × 50 μm device, $V_{GS} = 1.5$ V;$V_{DS} = 10$ V, (c) 2 × 100 μm device, $V_{GS} = -2$ V, $V_{DS} = 10$ V, and (d) 2 × 100 μm device, $V_{GS} = -1.5$ V, $V_{DS} = 10$ V.

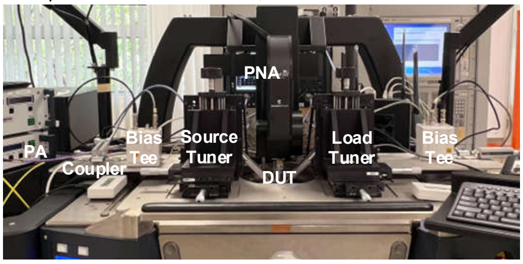


Fig. 10. Picture of load pull system.

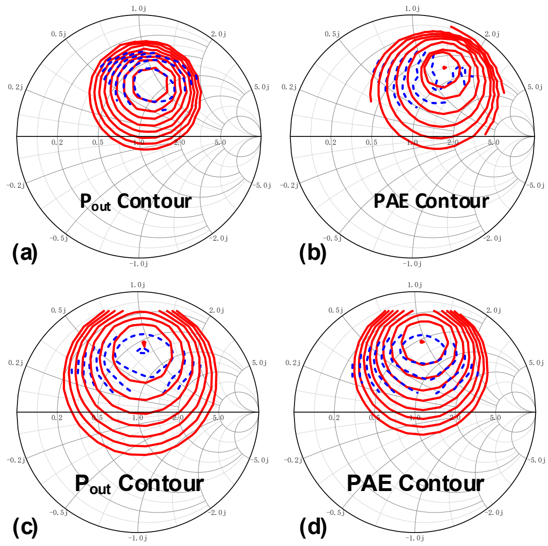


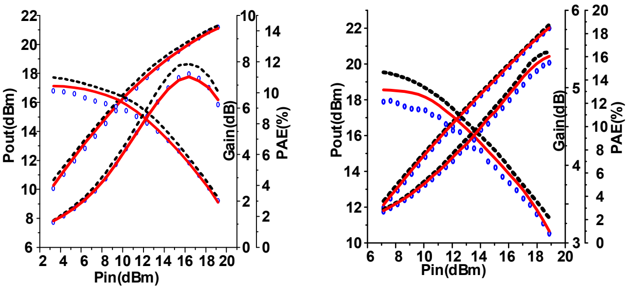
Fig. 11. Comparisons between measured (blue dashed lines) and simulated (red solid lines) loadpull contours for devices at 28.0 GHz. $V_{GS} = -2$ V and $V_{DS} = 10$ V.: (a) output power (P_{out}) contour, and (b) PAE contour of the 2 × 50 μm device, the input power is 10.9 dBm. The source reflection coefficient is 1.13 + j0.628. (c) output power (P_{out}) contour and (d) PAE contour of the 2 × 100 μm device, the input power is 15.7 dBm. The source reflection coefficient is 1.31 + j0.816.

technique, a novel GaN HEMT model was constructed within the ADS by employing symbolically define device. For comparative purposes, the conventional ASM-GaN-HEMT model was also instantiated to predict upon the empirical data. Notwithstanding the differences in handling of ψ_s and ψ_d both models-maintained congruity in other aspects, encompassing the extraction of direct current parameters and the modeling of both self-heating and trapping effects.

To corroborate the fidelity of the self-heating and trapping effect representations within the proposed model, DCIV characterizations were performed under different temperature conditions, alongside PIV characterizations under diverse non-zero quiescent bias states. Fig 8(a) and (b) illustrate measured versus modeled DCIV responses at disparate temperatures for the GaN HEMT with 2 × 50 μm gate width, thereby affirming the accuracy of the self-heating model. Fig 7 (c) demonstrates the alignment of measured and simulated PIV characteristics, which was influenced by different quiescent drain-source and gate-source voltages, indicating a satisfactory model of the trapping effects. Fig 8(d) contrasts the measured and modeled DCIV curves for the GaN HEMT with a 2 × 100 μm gate width, validating the model's scalability. Fig. 9 presents both the measured and simulated small-signal S-parameters for the two disparate GaN HEMT devices, revealing a strong concordance between simulation and measurement across a frequency range from 0.1 GHz to 40 GHz.

To substantiate the large-signal model delineated in this manuscript, a comprehensive load-pull characterization spanning the frequency range of 10 to 67 GHz was executed to assess the performance attributes of the GaN-based device. Fig. 10 provides a visual depiction of the load-pull platform. The platform employs a Keysight PNA-X N5247B network analyzer for signal generation, complemented by a Keysight B1500A unit supplying DC power for gate and source biasing. The tuners are the iCCMT-67100 from Focus Microwave, facilitating load and source tuning across the specified frequency spectrum. Power measurements were executed using a dual-channel power meter, the Keysight E4417A. The load-pull configuration further includes ancillary components such as input preamplifiers, broad-spectrum couplers, and bias tees, among others. Due to the bandwidth constraints of the coupler and isolator, which operate within the 26.5 GHz to 40 GHz range, this study presents large-signal test results exclusively at 28 GHz. Consequently, measurement and comparative analysis for harmonic and intermodulation distortion could not be conducted.

Fig. 11 presents the measured and simulated contours of the output power and Power Added Efficiency (PAE) derived from the load-pull analysis. These contours were acquired for two GaN HEMT devices with varying geometries, which are pertinent to 5G communication technologies. In Fig. 11 (a) and (b), the optimal output power (P_{out}) matching impedance is $0.409\angle 68°$ in simulation, compared to $0.433\angle 66.4°$ as observed in measurement. For the peak power-added efficiency (PAE), the optimal matching impedance is $0.612\angle 60.6°$ in simulation, whereas it measures at $0.7\angle 62.4°$. Fig. 11 (c) and (d) present a similar comparison: the best P_{out} matching impedance is $0.546\angle 84.7°$ in simulation and $0.58\angle 90°$ in



(a) (b)

Fig. 12. Measured (blue symbols), ASM model simulated (black dashed lines) and hybrid model simulated (red lines) output power (P_{out}), power gain (G), and PAE at 28.0 GHz. $V_{GS} = -2$ V and $V_{DS} = 10$ V. (a) 2×50 μm device, (b) 2×100 μm device.

measurement. Similarly, the best PAE matching impedance is $0.55\angle 86°$ in simulation and $0.562\angle 90°$ in measurement.

To further compare the large-signal performance of the theoretical model against the real-world device, a power sweep was conducted within the 10-67GHz loadpull architecture. Fig. 12 shows the measured (blue symbols), ASM model simulated (black dashed lines) and hybrid model simulated (red lines) values of power gain (G), output power (P_{out}), and PAE curves of the two transistors, although it seems not accurate in power gain, the error rate for each value itself is less than 5%. Therefore, a robust concordance between the measured data and simulation outputs could be observed.

V. Conclusion

In this work, we introduce an ANN-based modeling method for GaN HEMTs that integrates physical concepts. This framework is particularly tailored for the next generation of 5G power amplifiers. The modeling paradigm employs MOSPO to efficiently characterize the surface potentials. This ensures a globally accurate model that harmonizes the DC and RF characteristics based on the Advanced-SPICE-Model (ASM) theoretical constructs. The integration of ANN in the modeling process alleviates the complexities traditionally encountered in the ASM characterization of surface potentials, thereby retaining the integrity of the ASM's physical principles within the hybrid architecture. This fusion model surpasses the scalability limitations of conventional ANN-based models and demonstrates enhanced accuracy in predicting large-signal operations when benchmarked against standard ASM-derived physical models. The model also encompasses comprehensive treatments of self-heating and trapping phenomena, pivotal to the emulation of real-world device behavior. Developed within the ADS for a 100-nm GaN HEMT, the hybrid model's validity is evidenced by the close alignment of simulation results with empirical measurements, signifying the model's exceptional precision. The agreement between the theoretical predictions and measured data substantiates the efficacy of the proposed hybrid model, marking a significant stride in the nuanced field of GaN HEMT device modeling.

References

[1] A. Piacibello et al., “High-gain and high-linearity MMIC GaN Doherty power amplifier with 3-GHz bandwidth for Ka-Band satellite communications,” *IEEE Microwave and Wireless Technology Letters*, vol. 34, no. 6, pp. 765-768, June 2024.

[2] H. Xie et al., "GaN-on-Si HEMTs fabricated with Si CMOS-compatible metallization for power amplifiers in low-power mobile SoCs," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 2, pp. 141-144, Feb. 2021.

[3] H. Luo, W. Hu, and Y. Guo, “On large-signal modeling of GaN HEMTs: past, development and future,” *Chip*, vol. 2, no. 2, Art. no. 100052, Apr. 2023.

[4] Bottaro, Enrico, Santi Agatino Rizzo, and Nunzio Salerno. “Circuit models of power MOSFETs leading the way of GaN HEMT modelling—A review.” *Energies* vol. 15, no.9, Apr. 2022.

[5] M. Allaei, M. Shalchian and F. Jazaeri, “Modeling of short-Channel effects in GaN HEMTs,” *IEEE Transactions on Electron Devices*, vol. 67, no. 8, pp. 3088-3094, Aug. 2020, doi: 10.1109/TED.2020.3005122.

[6] Yang J, Jia Y, Ye N, et al., “A novel empirical IV model for GaN HEMTs,” *Solid-State Electronics*, vol. 14, no. 6, pp. 1-8, 2018.

[7] O. Jardel et al., “An electrothermal model for AlGaN/GaN power HEMTs including trapping effects to improve large-signal simulation results on high VSWR,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2660-2669, Dec. 2007.

[8] S. Aamir Ahsan, S. Ghosh, S. Khandelwal and Y. S. Chauhan, “Physics-based multi-bias RF large-signal GaN HEMT modeling and parameter extraction flow,” *IEEE Journal of the Electron Devices Society*, vol. 5, no. 5, pp. 310-319, Sept. 2017.

[9] U. Radhakrishna, P. Choi and D. A. Antoniadis, “Facilitation of GaN-based RF- and HV-bircuit designs using MVS-GaN HEMT compact model,” *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 95-105, Jan. 2019.

[10] S. Khandelwal et al., “ASM GaN: Industry standard model for GaN RF and power devices—Part 1: DC, CV, and RF Model,” *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 80-86, Jan. 2019.

[11] J. Xu, R. Jones, S. A. Harris, T. Nielsen and D. E. Root, “Dynamic FET model - DynaFET - for GaN transistors from NVNA active source injection measurements,” *2014 IEEE MTT-S International Microwave Symposium (IMS2014),* Tampa, FL, USA, 2014, pp. 1-3, doi: 10.1109/MWSYM.2014.6848293.

[12] Khusro A, Husain S, Hashmi M S, et al., “Small signal behavioral modeling technique of GaN high electron mobility transistor using artificial neural network: An accurate, fast, and reliable approach,” *International Journal of RF and Microwave Computer - Aided Engineering*, vol. 30, no. 4, Art. no. e22112, 2020, doi: 10.1002/mmce.22112

[13] H. Luo and Y. Guo, “A neural network-based modeling method for terminal charges in GaN HEMTs,” *2022 IEEE MTT-S International Wireless Symposium* (IWS), Harbin, China, pp. 1-3, 2022.

[14] A. Jarndal, A. Z. Markos, and G. Kompa, “Improved modeling of GaN HEMTs on Si substrate for design of RF power amplifiers,” *IEEE transactions on microwave theory and techniques*, vol. 59, no. 3, pp. 644-651, 2011.

[15] I. Angelov et al., “Large-signal modelling and comparison of AlGaN/GaN HEMTs and SiC MESFETs,” *2006 Asia-Pacific Microwave Conference*, pp. 279-282, 2006.

[16] U. Radhakrishna, S. Lim, P. Choi, T. Palacios, and D. Antoniadis, “GaNFET compact model for linking device physics, high voltage circuit design and technology optimization,” *2015 IEEE International Electron Devices Meeting* (IEDM), pp. 9.6. 1-9.6. 4, 2015.

[17] H. Luo, Z. Zhong, W. Hu, and Y. Guo, “Analysis and modeling of the temperature-dependent nonlinearity of intrinsic capacitances in AlGaN/GaN HEMTs,” *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 4, pp. 373–376, Apr. 2021.

[18] X. Du et al., “ANN-based large-signal model of AlGaN/GaN HEMTs with accurate buffer-related trapping effects characterization,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 7, pp. 3090-3099, July 2020, doi: 10.1109/TMTT.2020.2990171.

[19] W. Hu and Y. -X. Guo, “An Evolutionary Multilayer Perceptron-Based Large-Signal Model of GaN HEMTs Including Self-Heating and Trapping Effects,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 2, pp. 1146-1156, Feb. 2022.

[20] W. Hu, H. Luo, X. Yan and Y. -X. Guo, "An accurate neural network-based consistent gate charge model for GaN HEMTs by refining intrinsic capacitances," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 7, pp. 3208-3218, July 2021.

[21] A.-D. Huang, Z. Zhong, W. Wu, and Y.-X. Guo, “An artificial neural network-based electrothermal model for GaN HEMTs with dynamic trapping effects consideration,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 8, pp. 2519-2528, 2016.

[22] A. Jarndal, S. Husain, M. Hashmi, “On temperature-dependent small-signal modelling of GaN HEMTs using artificial neural networks and support vector regression,” *IET Microwaves, Antennas & Propagation,* vol. 15, no. 8, pp. 937-953, 2021.

[23] A. Jarndal, S. Husain, 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*, 2021, 31(3): e22542.

[24] H. Luo, Z. Zhong, W. Hu and Y. Guo, “Analysis and modeling of the temperature-dependent nonlinearity of intrinsic capacitances in AlGaN/GaN HEMTs,” *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 4, pp. 373-376, April 2021.

[25] A. Jarndal, S. Husain, M. Hashmi and F. M. Ghannouchi, “Large-signal modeling of GaN HEMTs using hybrid GA-ANN, PSO-SVR, and GPR-Based approaches,” *IEEE Journal of the Electron Devices Society*, vol. 9, pp. 195-208, 2021,

[26] Mishra, Ashutosh, et al., “Multilayer perceptron–random forest-based hybrid machine learning–neural network model for GaN high electron mobility transistor's parameter estimations,” *International Journal of RF and Microwave Computer-Aided Engineering,* 2022, 32(7): e23191.

[27] W. Liu, W. Na, L. Zhu, J. Ma and Q. -J. Zhang, "A Wiener-type Dynamic neural network approach to the modeling of nonlinear microwave devices," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 6, pp. 2043-2062, June 2017.

[28] 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, Nov. 2022.

[29] Khandelwal S, Fjeldly T A., “A surface-potential-based compact model for study of non-linearities in AlGaAs/GaAs HEMTs,” *2012 IEEE Compound Semiconductor Integrated Circuit Symposium* (CSICS). pp. 1-4, 2012.

[30] S. Khandelwal et al., “Robust surface-potential-based compact model for GaN HEMT IC design,” *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3216-3222, Oct. 2013.

[31] S. Aamir Ahsan, A. -u. -H. Pampori, S. Ghosh, S. Khandelwal and Y. S. Chauhan, “A new small-signal parameter extraction technique for large gate-periphery GaN HEMTs,” *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 10, pp. 918-920, Oct. 2017.

[32] S. Khandelwal, N. Goyal, and T. A. Fjeldly, “A physics-based analytical model for 2DEG charge density in AlGaN/GaN HEMT devices,” *IEEE Trans. Electron Devices*, vol. 58, no. 10, pp. 3622–3625, Oct. 2011.

[33] S. Khandelwal and T. A. Fjeldly, “A physics based compact model of I–V and C–V characteristics in AlGaN/GaN HEMT devices,” *Solid-State Electron.*, vol. 76, pp. 60–66, Oct. 2012.

[34] Reyes-Sierra M, Coello C A C. “Multi-objective particle swarm optimizers: A survey of the state-of-the-art,” *International journal of computational intelligence research*, vol. 2, no. 3, pp. 287-308, 2006.

[35] Z. Lu, H. Xie, J. Piao, W. Zhengzhe, N. G. Ing and Y. Zheng, “A wideband GaN HEMT modelling with comprehensive hybrid parameter extraction for 5G power amplifiers,” *2023 IEEE International Symposium on Circuits and Systems* (ISCAS), Monterey, CA, USA, pp. 1-5, 2023.

[36] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, “A new method for determining the FET small-signal equivalent circuit,” *IEEE Trans. Microw. Theory Techn.*, vol. 36, no. 7, pp. 1151–1159, Jul. 1988.

[37] R. G. Brady, C. H. Oxley, and T. J. Brazil, “An improved small-signal parameter-extraction algorithm for GaN HEMT devices,” *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 7, pp. 1535–1544, Jul. 2008.

[38] A. Alt, D. Marti, and C. R. Bolognesi, “Transistor modeling: Robust small-signal equivalent circuit extraction in various HEMT technologies,” *IEEE Microw. Mag.*, vol. 14, no. 4, pp. 83–101, Jun. 2013.

[39] J. Heaton, "Introduction to Neural Networks for Java," Heaton Research, 2008.

[40] A. B. Nassif, D. Ho, and L. F. Capretz, "Towards an early software estimation using log-linear regression and a multilayer perceptron model," *Journal of Systems and Software*, vol. 86, no. 1, pp. 144-160, 2013.

[41] S. García-Sánchez, I. Íñiguez-de-la-Torre, T. González, and J. Mateos, "Hybrid AI-Thermal Model Trained via Monte Carlo Simulations to Study Self-Heating Effects," IEEE Transactions on Electron Devices, 2024.

[42] O. Jardel et al., “An Electrothermal Model for AlGaN/GaN Power HEMTs Including Trapping Effects to Improve Large-Signal Simulation Results on High VSWR,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 12, pp. 2660-2669, Dec. 2007.