An Efficient Transistor Modelling Method using Convolutional Neural Network with Transfer Learning

***Abstract*— In this paper, an efficient CNN model with transfer learning ability is developed for predicting the nonlinear behavior of transistors under large signal stimuli. The CNN consists of multiple hierarchical layers where different levels capture different nonlinear relationships. Meanwhile, we propose fine-tuning neural layers to minimize the number of measurement data under varying operational conditions of transistors. A 0.25 um GaN HEMT transistor was selected as the device under test for this experiment. Using the proposed transfer learning based method, only 30% of the conventional measurement data is needed to achieve a normalized mean squared error (NMSE) below -40 dB. This proposed approach provides a practical solution for efficient transistor behavioral models.**

I. Introduction

Radio frequency power amplifiers (RFPAs), of which the RF power transistor is the essential part [1], are vital components in modern wireless communication. Transistor models play a very important role in understanding and predicting transistor behavior, thus guiding circuit design. Transistors have traditionally been characterized by physical basis models or empirical basis models. However, with the introduction of new materials and the complicated application scenarios, the physical characteristics of transistors have become considerably more complex, making accurate modelling becomes a challenging or even impossible task [2].

Recently, neural network-based modeling methods have gained widespread adoption in the modeling of microwave devices, primarily due to their rapid prototyping ability with high accuracy. [3]. However, the model development cycle is still constrained by the extensive need for nonlinear measurements to gather data for model training. To tackle this challenge, many prevailing strategies have been employed, which includes optimizing sampling algorithms or utilizing knowledge-based neural networks (KBNN). One sampling algorithm selects informative samples across the entire design space to reduce training samples [4]. However, its effectiveness depends on design space complexity. KBNN integrates neural networks with prior knowledge to create empirical circuit models [5], decreasing training data and improving extrapolation performance. But KBNN approaches require a physical/compact model as as starting point and also suffers the lacking of measurement data.

This paper presents an efficient CNN modeling approach that combines existing transistor neural network models with new transistor measurement data. It employs the fine-tuning of convolutional neural networks to reduce the required sample size for neural network modeling. This research validates the CNN model's precise prediction of the transmitted wave of GaN HEMT through large signal load-pull experiments across various working frequencies and DC biases.

II. RF Transistor and CNN Model

A. *Transistor Modeling*

Generally, both the RF transistor and its model can be treated as a two-port network to describe its input-output characteristics, as shown in Fig. 1. The NN behavioral model of the transistor aims to obtain the transmission wave b_2 by known inputs a_1 and a_2 [6], so that we can calculate the transistor's gain, PAE, and other performance indicators.

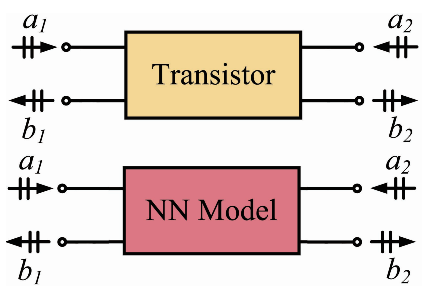


Fig. 1. Two-port network of Transistor and Model with incident and reflected power waves.

Since the transistor's scattered waves contain phase information, the numerical forms are complex. Consequently, we represent scattered waves by real and imaginary parts, For example, b_2 can be represented as a function of a_1 and a_2:

$$(b_2^R, b_2^I) = F(a_1^R, a_1^I, a_2^R, a_2^I) \tag{1}$$

B. *Transfer Learning*

Transfer learning is a promising machine learning methodology for solving the problem of limited training data. The concept of transfer learning has been successfully applied in many applications since proposed [7], [8]. As shown in Fig. 2, the Pre-model is trained on the source domain, denoted as D^S, where training samples are collected by measuring the transistor under certain working conditions. In contrast, new measurement data is obtained from the target domain, denoted as D^T, which represents the transistor operating under target working conditions. Given corresponding learning task, transfer learning strives to improve the learning in D^T using the knowledge learned in D^S, where $D^S \neq D^T$.

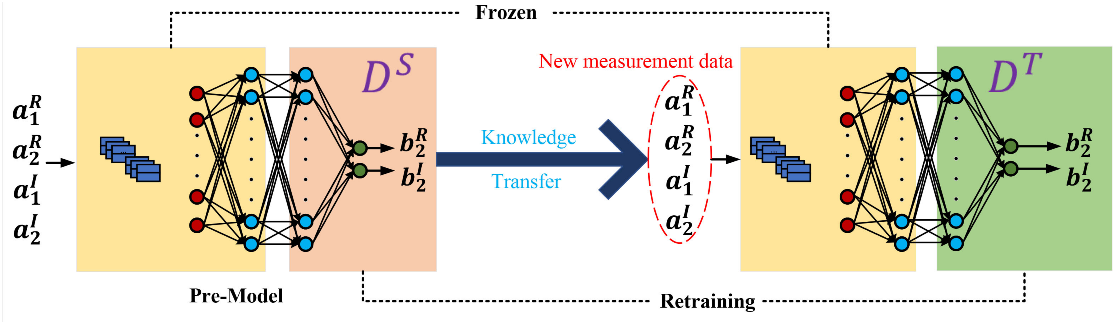


Fig. 2. Proposed transfer learning process for transistor modeling.

C. Proposed CNN Based Transfer Learning Approach

Fine-tuning is a common method in transfer learning. As shown in Fig. 2, the yellow region represents frozen layers where the parameters remain unchanged during the transfer process, the source domain model is initially trained, followed by updating the fine-tuning layers of the model using the data in target domain, resulting in the target domain model.

Let X^S, Y^S, X^T and Y^T represent the input and output from two domains, respectively. The number of samples in the target domain is limited to 30% of that in the source domain, which is reflected in subsequent impedance points selection. Initialize the weight parameters of the Pre-model as W^S and define the loss function for the Pre-model as [9]:

$$L^S = \frac{1}{N}\sum_{i=1}^{N}|(f(W^S, X^S) - Y^S)| \tag{2}$$

where N represents the sample size and f represents the neural network's forward propagation to get predictions. W^S can be updated by back-propagation algorithm.

After obtaining the Pre-model, enter the new measurement data into the Pre-model, update the weight parameters of the fine-tuning layers, and the transfer model can be obtained according to:

$$L^T = \frac{1}{n}\sum_{i=1}^{n}|(f(W^S, X^T) - Y^T)| \tag{3}$$

where n represents the sample size in the D^T.

Due to W^S and W^T sharing same features, the required (X^T, Y^T) is considerably reduced. The accuracy of the transistor model can be expressed in form of normalized mean squared error (NMSE) in dB:

$$NMSE_{dB} = 10log_{10}\frac{\sum[\hat{Y}(i) - Y(i)]^2}{\sum(Y(i))^2} \tag{4}$$

where $\hat{Y}(i)$ and $Y(i)$ represent the predicted values and the measured values.

III. Model Validation

A. Getting Data

In order to obtain the training data required as shown in Fig. 2, NVNA with active load-pull test is performed on the Device Under Test(DUT). Fig. 3 illustrates the block diagram of the active load-test test. A Keysight-N5247 PNA-X is taken as the main receiver of the test system to receive a_1, a_2, b_1, b_2 sent back by the couplers. Fig. 4 (a) shows a front photo of the active load-pull test setup. The GaN device under the microscope is shown in Fig. 4 (b). The measurement data within the D^S comes from condition A (note: 8 GHz, VGS bias=-2 V, VDS bias=18 V), and measurement data within the D^T is derived from condition B (note: 12 GHz, VGS bias=-2.4 V, VDS bias=15 V).

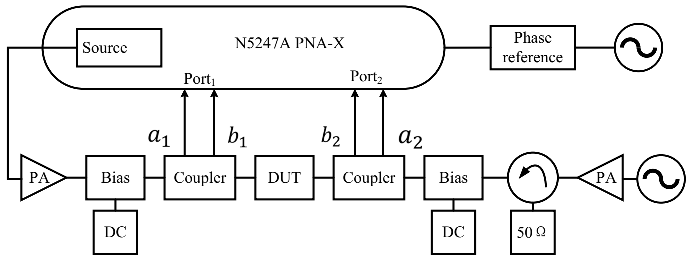
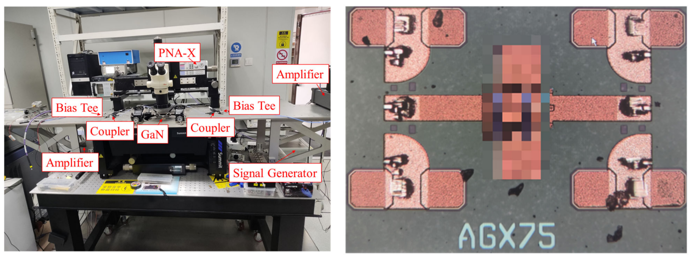


Fig. 3. Real-time On-chip Load-pull measurement system.



(a) (b)

Fig. 4. (a) Photograph of measurement system; (b) The GaN device.

B. Pre-model

During the measurement data acquisition phase in the D^S, we first performed under condition A with a termination load of 50 ohms. The input power ranged from 7 dBm to 22 dBm, with a step size of 0.5 dBm as shown in Fig. 5 (a). Subsequently, the determination of 73 impedance points for load sweeping under all input power was carried out as shown in Fig. 5 (b), which are in the region of from optimum output power impedance to 2.5 dB back-off. Therefore, the

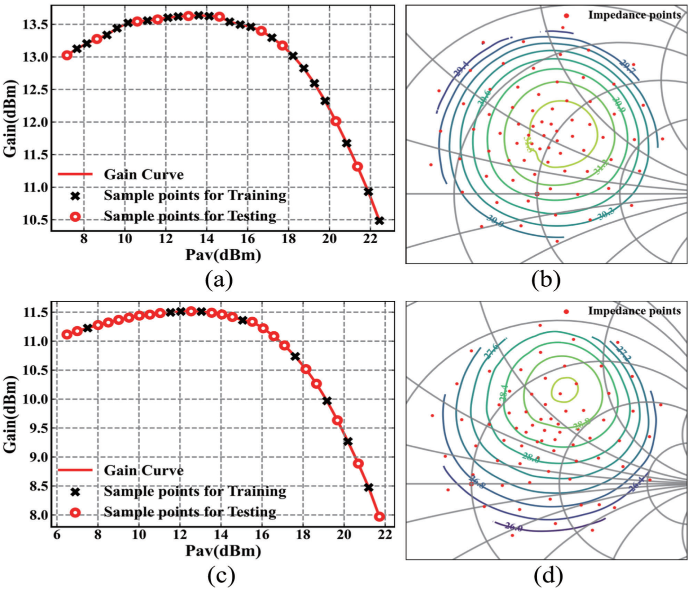


Fig. 5. (a) Power sweep curves of the DUT in Pre-model experiment; (b) load-pull impedance points used in Pre-model experiment; (c) Power sweep curves of the DUT in transfer model experiment; (d) load-pull impedance points used in transfer model experiment.

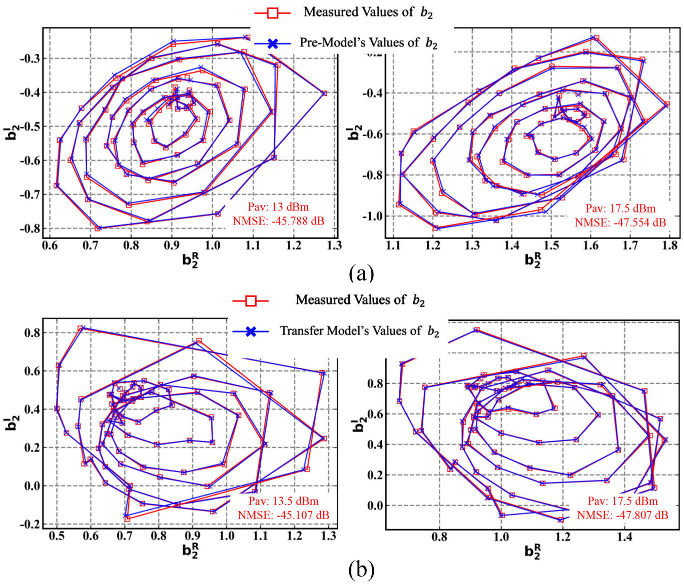


Fig. 6. (a) Measured and Pre-model's values of b_2 and the corresponding input power; (b) Measured and transfer model's values of b_2 and the corresponding input power.

D^S measurement data has a total of 2263 sets. Approximately 70% of the measurement data is utilized for training, and the remaining 30% is for validation. Fig. 6 (a) displays the comparison between Pre-model's and measured values of b_2 for input power of 13 dBm and 17.5 dBm.

C. Transfer Model

Once the Pre-model has been acquired, the prediction ability over other frequencies and biases needs to be developed. Fig. 5 (c) and (d) illustrate the power sweep and load impedance points under condition B, involving also 2263 sets. It can be observed that the sample points for training have been significantly reduced, comprising only 30% of the total measurement data. Fig. 6 (b) illustrates the comparison between the predicted and measured values of b_2 at input powers of 13.5 dBm and 17.5 dBm in the transfer model test dataset. By comparing Fig. 6 (a) and Fig. 6 (b), it is evident that despite a substantial reduction in sample points for training, the transfer model's NMSE has not significantly decreased, demonstrating the successful knowledge transfer from the Pre-model to the transfer model.

IV. Conclusion

Transfer learning has been effectively implemented to propose a robust, fast, and accurate model for large-signal transistor models in this work. With only 30% of the typically required measurement quantity, transistor performance under different conditions can be extrapolated. This method has been verified using measurement data from a real device, and the $NMSE_{dB}$ of the B_{21} wave is below -40 dB for all data points. The proposed approach promises a generic new solution for a more effective transistor model, especially with the limited data and time provided.

References

[1] J. J. Komiak, "GaN HEMT: Dominant Force in High-Frequency Solid-State Power Amplifiers," *IEEE Microwave Magazine*, vol. 16, no. 3, pp. 97–105, 2015.

[2] L. Dunleavy, C. Baylis, W. Curtice, and R. Connick, "Modeling GaN: Powerful but Challenging," *IEEE Microwave Magazine*, vol. 11, no. 6, pp. 82–96, 2010.

[3] D. E. Root, "Future device modeling trends," *IEEE Microwave Magazine*, vol. 13, no. 7, pp. 45–59, 2012.

[4] J. Wei, H. Wang, T. Zhao, Y.-L. Jiang, and J. Wan, "A new compact mosfet model based on artificial neural network with unique data preprocessing and sampling techniques," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 4, pp. 1250–1254, 2023.

[5] F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial neural networks for microwave computer-aided design: The state of the art," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4597–4619, 2022.

[6] J. Cai, J. Wang, C. Yu, H. Lu, J. Liu, and L. Sun, "An artificial neural network based nonlinear behavioral model for RF power transistors," in *2017 IEEE Asia Pacific Microwave Conference (APMC)*, 2017, pp. 600–603.

[7] H. Wang, K. Wang, J. Yang, L. Shen, N. Sun, H.-S. Lee, and S. Han, "GCN-RL Circuit Designer: Transferable Transistor Sizing with Graph Neural Networks and Reinforcement Learning," in *2020 57th ACM/IEEE Design Automation Conference (DAC)*, 2020, pp. 1–6.

[8] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.

[9] A. M. Taqi, A. Awad, F. Al-Azzo, and M. Milanova, "The Impact of Multi-Optimizers and Data Augmentation on TensorFlow Convolutional Neural Network Performance," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2018, pp. 140–145.