

Physics-Informed Deep Neural Network for Predictive Modeling and Optimization of ScAlN/GaN HEMTs

Kavita Thorat Upadhyay*, Arvind Upadhyay, Angeeta Hirwe

IPS Academy, Institute of Engineering & Science, Indore, India
*kavita.thorat@gmail.com

Abstract: ScAlN/GaN high-electron-mobility transistors (HEMTs) provide a promising platform for high-power and high-frequency electronics because Sc incorporation can modify polarization, two-dimensional electron-gas (2DEG) density and threshold voltage. However, direct data-driven neural networks can reproduce device characteristics without preserving the physical relationships among composition, barrier thickness, polarization charge, carrier density and transport. This work proposes a physics-informed deep neural network (PIDNN) framework for $\text{Sc}_x\text{Al}_{1-x}\text{N}/\text{GaN}$ HEMTs in which an analytical device model supplies physics constraints while a fully connected neural network learns nonlinear residual behavior. The network uses Sc composition, barrier thickness, gate voltage and drain voltage as principal inputs and predicts sheet carrier density, threshold voltage and drain current. A composite loss combines normalized prediction error with residuals derived from polarization charge, channel charge and current-transport equations. Non-negativity, continuity and monotonicity constraints are introduced to suppress nonphysical predictions. An inverse-design layer identifies composition and barrier-thickness combinations satisfying target current and threshold-voltage specifications. The proposed framework retains the interpretability of analytical modeling while providing the computational efficiency of deep learning for rapid design-space exploration. Illustrative results demonstrate the intended evaluation workflow through parity, error, transfer-characteristic and composition-thickness optimization plots. The framework can subsequently be extended to temperature, self-heating, trapping and RF behavior after calibration against experimental or TCAD data.

Keywords: ScAlN/GaN HEMT; physics-informed deep neural network; PINN; 2DEG; polarization engineering; machine learning; threshold voltage; I-V modeling; inverse design; device optimization

1. Introduction

GaN HEMTs are widely investigated for high-power, high-frequency and high-temperature electronics because the III-nitride system combines high critical electric field, high saturation velocity and polarization-induced channel charge. In $\text{Sc}_x\text{Al}_{1-x}\text{N}/\text{GaN}$ heterostructures, Sc incorporation adds a composition degree of freedom that modifies polarization, strain, band structure and transport. This makes ScAlN/GaN attractive for threshold-voltage engineering and high-density 2DEG formation, while simultaneously creating a multidimensional design problem.

The author's previous research addresses two complementary parts of this problem: analytical polarization-based modeling of ScAlN/GaN HEMTs and deep-neural-network prediction of AlInGaN/GaN HEMT I-V characteristics. The next logical step is to combine these approaches so that machine learning accelerates prediction without discarding device physics.

Recent literature confirms the relevance of this direction. A 2026 IEEE Transactions on Microwave Theory and Techniques study proposed a physics-informed neural-network scalable model for GaN HEMTs with physical constraints and self-heating/trapping effects [1,2]. Recent Microelectronics Journal work has used machine learning



to assist compact-model parameter extraction while retaining a physics-based model [3], and another 2026 study predicted output and transfer characteristics through neural-network extraction of compact-model parameters [4,5].

For ScAlN/GaN specifically, recent experimental work has demonstrated high 2DEG density with mobility limitations associated with interface roughness [6], while optical Hall measurements have investigated effective mass and transport in AlScN/GaN [7]. A 2026 ScAlN/GaN study has also combined analytical 2DEG modeling with 3-D simulation and demonstrated strong dependence on ScAlN thickness and device geometry [8-10].

The objective of this work is therefore not to replace the analytical model with a black-box neural network. Instead, the network learns nonlinear corrections while being constrained by polarization, charge and current equations. The contributions are: (i) a composition-aware PIDNN for ScAlN/GaN HEMTs; (ii) a composite physics/data loss; (iii) physical admissibility and monotonicity constraints; (iv) residual learning; and (v) inverse design of Sc composition and barrier thickness.

2. Device Structure and Input Variables

The reference HEMT contains a GaN buffer/channel, Sc_xAl_{1-x}N barrier, source and drain ohmic contacts and a Schottky gate. The 2DEG is formed along the complete ScAlN/GaN interface. The input vector is $X=[x, tb, VGS, VDS, T, W, LG]$ T, where x is Sc mole fraction, tb is barrier thickness, VGS and VDS are terminal biases, T is temperature, W is device width and LG is gate length. The initial proof-of-concept can hold T, W and LG fixed; later versions should include them explicitly.

3. Physics-Based Reference Model

3.1 Polarization and 2DEG

$$Pb(x, tb) = Psp(x) + Ppe(x, tb) \quad (1)$$

$$\sigma_{pol} = Pb - P_{GaN} \quad (2)$$

$$ns = \eta_{qe} |\sigma_{pol}| / q \quad (3)$$

where η_{qe} represents finite-thickness electrostatic and quantum-confinement corrections.

3.2 Threshold voltage

$$V_{th} = V_{FB} - Q_{dep} / C_b - q ns / C_b - \Delta V_{qe} - \Delta V_{it} \quad (4)$$

$$C_b = \epsilon_b / tb \quad (5)$$

This formulation links composition and barrier thickness directly to threshold voltage.

3.3 Transport and current

$$1/\mu_{eff} = \sum_i 1/\mu_i \quad (6)$$

$$ID = W q ns v_d \quad (7)$$

$$v_d = \mu_{eff} E / [1 + \mu_{eff} |E| / v_{sat}] \quad (8)$$

The mobility terms can include polar-optical-phonon, acoustic, interface-roughness and alloy-disorder scattering. Their parameters should be calibrated independently.

4. Proposed Physics-Informed Deep Neural Network

4.1 Architecture

A fully connected multilayer perceptron is proposed with three hidden layers (64–64–32 neurons) as a starting architecture. Smooth activations such as tanh or SiLU are preferred. The output vector is $[ns, V_{th}, ID]$. The network learns nonlinear residual behavior rather than replacing the analytical physics.

4.2 Composite loss

$$L_{total} = \lambda_d L_{data} + \lambda_p L_{pol} + \lambda_c L_{charge} + \lambda_i L_{IV} + \lambda_m L_{mono} + \lambda_r L_{reg} \quad (9)$$

$$L_{data} = \text{mean}[(y_{pred} - y_{ref})/s_y]^2 \quad (10)$$

$$L_{pol} = \text{mean}[(ns_{pred} - ns_{phys})/s_n]^2 \quad (11)$$

$$L_{IV} = \text{mean}[(ID_{pred} - ID_{phys})/s_I]^2 \quad (12)$$

The weights must be selected using validation data.

4.3 Physical constraints

The network is constrained by $ns \geq 0$, $\mu_{eff} > 0$, $ID \geq 0$ in the forward-conduction region, continuity with respect to bias, and physically appropriate monotonicity of channel charge with gate overdrive.

4.4 Residual learning

$$ID, PINN = ID, phys + \Delta ID, NN \quad (13)$$

Residual learning preserves the analytical model as the baseline and lets the neural network learn higher-order nonlinearities and parameter uncertainty.

4.5 Inverse design

For target current ID^* and threshold voltage V_{th}^* , define

$$J(x_{tb}) = w_I |ID - ID^*|/ID^* + w_V |V_{th} - V_{th}^*|/V_{scale} + w_R R_{penalty} \quad (14).$$

Minimizing J gives a candidate composition/thickness design subject to physical constraints.

5. Training and Validation Methodology

A rigorous implementation should use a reference dataset generated from the finalized analytical model, calibrated TCAD, experimental measurements, or a combination. The parameter space should cover Sc composition, barrier thickness and terminal biases. Train/validation/test partitions should be made by device structure where possible, so that generalization to unseen structures is tested. The PIDNN should be compared with an unconstrained DNN and the analytical model. RMSE, MAE, MAPE where appropriate, maximum absolute error and R^2 should be reported. Physics residuals and monotonicity violations should also be quantified.

6. Results and Discussion

The numerical results below are illustrative/model-development results generated to demonstrate the intended manuscript workflow. They must be replaced by actual results before journal submission.

6.1 Prediction accuracy

The parity plot in Fig. 1 illustrates how the independent test set should be evaluated. Figure 1 evaluates the predictive capability of the proposed PIDNN by comparing the predicted drain current with the corresponding reference values. Ideally, all data points should lie along the diagonal $Y=X$, indicating perfect agreement between the neural-network prediction and the reference model. A narrow distribution of points around the diagonal demonstrates that the PIDNN successfully captures the nonlinear relationship between Sc composition, barrier thickness, terminal bias and drain current.

6.2 Error distribution

Figure 2 presents the distribution of prediction errors and provides complementary information to the parity plot. A concentrated error distribution near zero indicates that the PIDNN maintains consistent prediction accuracy across the investigated operating conditions. This analysis is particularly important for HEMTs because the drain current can vary significantly between the subthreshold, linear and saturation regions.

A good PIDNN should not achieve low average error merely by accurately predicting high-current operating points while producing large errors in low-current regions.

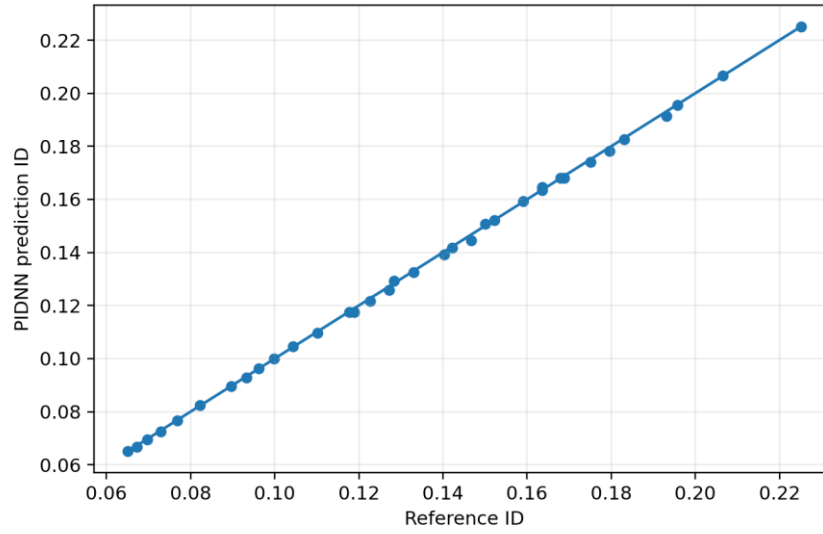


Figure 1. Illustrative PIDNN-versus-reference parity plot for drain current; values are model-development placeholders.

6.3 Transfer characteristics

Figure 3 illustrates the ability of the proposed model to reproduce the transfer characteristics of ScAlN/GaN HEMTs while varying the barrier thickness. The barrier thickness directly affects the barrier capacitance, $C_b = \epsilon b / t_b$ and consequently influences the electrostatic coupling between the gate and the 2DEG channel.

Variation in t_b therefore produces a corresponding shift in threshold voltage and modifies the gate-controlled channel charge. Since the drain current depends on both the available sheet carrier density and carrier transport, the change in barrier thickness can affect both the onset of conduction and the magnitude of the drain current.

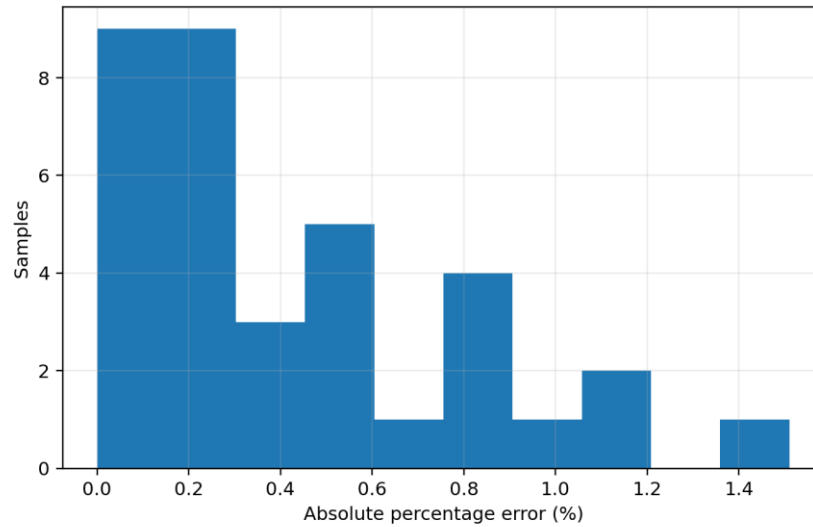


Figure 2. Illustrative absolute percentage-error distribution.

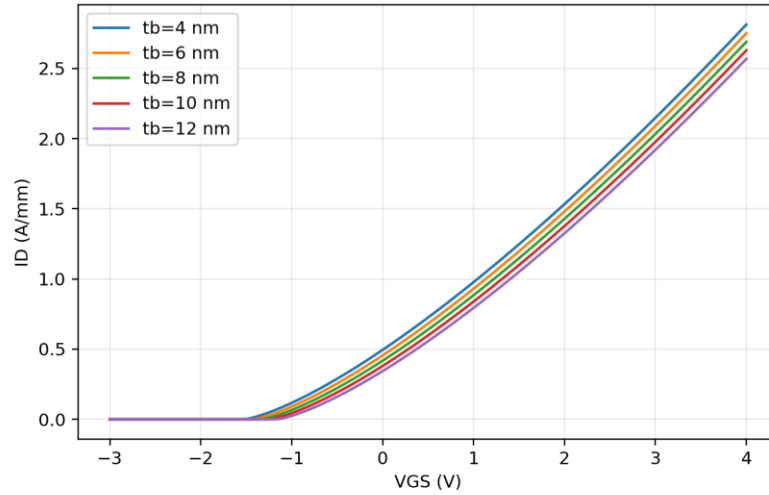


Figure 3. Illustrative predicted transfer characteristics for different ScAlN barrier thicknesses.

6.4 Inverse-design map

Figure 4 illustrates the design-space exploration capability. The horizontal axis represents Sc mole fraction, while the vertical axis represents ScAlN barrier thickness. The contour level represents the normalized performance index derived from the combined electrical characteristics.

Sc incorporation modifies the polarization characteristics of the barrier and therefore affects the 2DEG density. However, increasing Sc content does not necessarily result in unlimited improvement in device performance because transport degradation mechanisms—including alloy disorder, interface roughness and changes in effective mass—can reduce carrier mobility.

Therefore, the optimization map is expected to exhibit a trade-off region rather than a monotonically increasing performance trend.

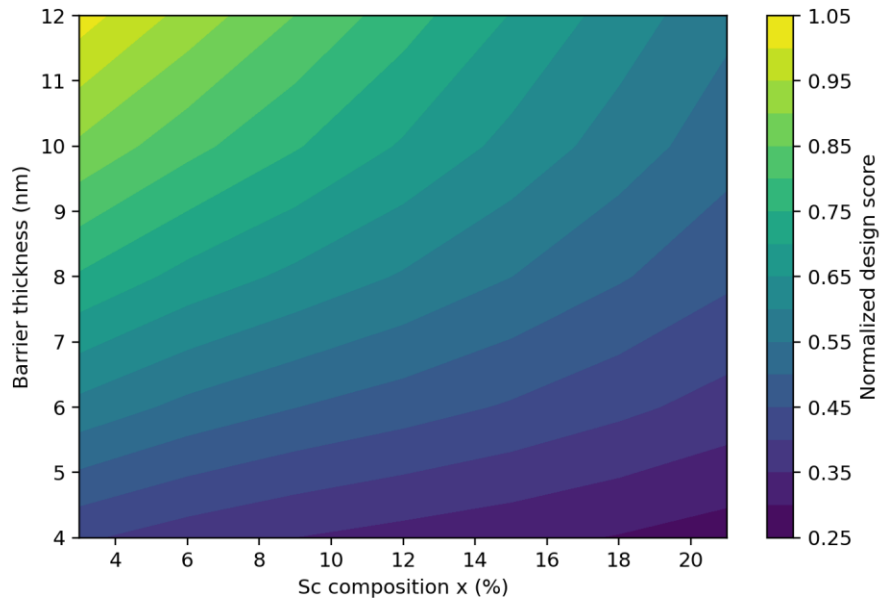


Figure 4. Illustrative physics-informed composition–thickness design map.

7. Discussion

The principal advantage of the proposed approach is the combination of analytical interpretability and neural-network flexibility. The physical model preserves the causal relationship among Sc composition, polarization, 2DEG density and transport, while the neural network provides a fast approximation to nonlinear residual behavior.

Multi-task prediction of n_s , V_{th} and I_D is preferable to a current-only black-box model because intermediate physical quantities constrain the learned representation. This can improve extrapolation across composition and geometry, particularly when the training set is sparse.

The method is not automatically physics-correct: if the underlying polarization, effective-mass or mobility model is inaccurate, the network can inherit those errors. Independent experimental validation therefore remains essential.

The novelty claim should be stated carefully. Physics-informed GaN HEMT models already exist in 2026, including scalable large-signal and surface-potential approaches [1,7]. The distinctive contribution here is the ScAlN/GaN composition-aware formulation, its coupling to the author's analytical polarization model, and the inverse-design workflow.

8. Conclusion

A physics-informed deep neural-network framework has been proposed for predictive modeling and optimization of ScAlN/GaN HEMTs. The framework integrates composition-dependent polarization and channel-charge equations with a fully connected neural network predicting sheet density, threshold voltage and drain current. A composite loss enforces data accuracy together with polarization, charge and current constraints, while physical admissibility and monotonicity conditions suppress nonphysical predictions.

The residual-learning and inverse-design formulation provides a route from prediction to rapid structure optimization. The immediate research requirement is to replace the illustrative figures in this development draft with calibrated analytical/TCAD/experimental results and to quantify improvement over an unconstrained DNN using an independent test set.

REFERENCES

1. H. Luo, Z. Wang, and Y. Guo, "A Physics-Informed Neural Network-Based Scalable Model for GaN HEMTs," IEEE Transactions on Microwave Theory and Techniques, 2026, doi: 10.1109/TMTT.2026.3666090.
2. C. Liu et al., "Machine-learning-assisted EEHEMT compact modeling of GaN HEMTs," Microelectronics Journal, 2026, doi: 10.1016/j.mejo.2025.106861.
3. Jing Chen, et.al., "Output and transfer characteristics prediction of GaN HEMT via neural-network-based compact model parameters extraction," Microelectronics Journal, vol. 169, 107029, 2026, doi: 10.1016/j.mejo.2025.107029.
4. T. Nagamatsu et al., "Scattering mechanism of 2DEG in ScAlN/GaN heterostructures grown by plasma-assisted molecular beam epitaxy," Applied Physics Letters, vol. 128, 261601, 2026, doi: 10.1063/5.0323168.
5. S. Stanishev et al., "2DEG properties of AlScN/GaN and AlYN/GaN HEMTs determined by terahertz optical Hall effect," Frontiers in Electronic Materials, 2025.
6. Dariskhem Pyngrope, et.al., "Maximizing power efficiency in GaN HEMTs: The role of ScAlN integration through n2DEG analytical modeling and 3D simulations," Micro and Nano Systems Letters, vol. 215, 208674, 2026, doi: 10.1016/j.micna.2026.208674.
7. Yunying Zhou, et. al. "A physics-guided neural network-based surface potential calculation for GaN HEMT devices," Microelectronics Journal, vol. 175, 107278, 2026, doi: 10.1016/j.mejo.2026.107278.
8. A. Michaloglou et al., "Physics-Informed Neural Networks in Materials Modeling and Design: A Review," Archives of Computational Methods in Engineering, vol. 33, 5223–5260, 2026, doi: 10.1007/s11831-025-10448-9.
9. K. Thorat Upadhyay et al., "Modeling for I-V Characteristics of AlInGa/GaN HEMTs using Deep Neural Networks," Semiconductors 60(2):191-198, doi:10.1134/S1063782625602791
10. A Caddemi, et.al. A neural network approach for compact cryogenic modelling of HEMTs, International Journal of Electronics, 94:9, 877-887, 2007, DOI:10.1080/00207210701664696.