Journal of Applied Sciences and Modelling

https://journal.krrishpub.com/index.php/jasm/index

Modeling Advanced Electronic Device Characterization Using a Unified Physics-Informed Machine Learning Framework

Ebrahim E. Elsayed [a,*], Mohammed R. Hayal [a], Davron Aslonqulovich Juraev [b,c,d], Jo'shqin Shakirovich Abdullayev [e]

[a] *Department of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt.*
[b] *Scientific Research Center, Baku Engineering University, Baku, AZ0102, Azerbaijan.*
[c] *Department of Scientific Research, Innovation and Training of Scientific and Pedagogical Staff, University of Economics and Pedagogy, Karshi, 180100, Uzbekistan.*
[d] *Department of Mathematical Analysis and Differential Equations, Karshi State University, Karshi, 180119, Uzbekistan.*
[e] *Department of Physics and Chemistry, TIIAME National Research University (The Tashkent Institute of Irrigation and Agricultural Mechanization Engineers), Tashkent, 100000, Uzbekistan.*

Abstract

Advancements in nanoscale semiconductor technologies, wide-bandgap materials, and emerging 2D devices have significantly increased the complexity of device characterization, making traditional TCAD simulations and compact modeling increasingly time-consuming and insufficiently flexible. Physics-Informed Machine Learning (PIML) offers a promising pathway by integrating physical laws with data-driven modeling to enhance accuracy, interpretability, and generalization. This paper presents a unified PIML framework for advanced electronic device characterization that embeds semiconductor transport physics, electrostatic constraints, and compact-model priors directly into neural architectures and loss functions. The proposed approach leverages hybrid PINN-based solvers, residual-learning compact models, and uncertainty-aware training to model I–V, C–V, and RF characteristics across nanosheet FETs, GaN HEMTs, and 2D-material THz transistors. Experimental results demonstrate that PIML models reduce prediction error by up to 35% compared to purely data-driven models and achieve substantially improved physical consistency, particularly in charge conservation and monotonic device behavior. Moreover, the differentiable structure of the framework enables efficient inverse design and parameter extraction, significantly accelerating device optimization workflows. Overall, the study establishes physics-informed ML as a scalable and robust methodology for next-generation electronic device modeling, bridging the gap between high-fidelity physics solvers and fast, design-oriented surrogate models.

Keywords: Physics-Informed Machine Learning (PIML), Semiconductor device modeling, Physics-Informed Neural Networks (PINNs), Compact models, TCAD surrogate modeling, Nanosheet FETs.

Article Information:
DOI: https://doi.org/10.71426/jasm.v1.i1.pp9-18
Received: *29 Nov. 2025* | Revised: *29 Dec. 2025* | Accepted: *29 Dec. 2025* | Published: *31 Dec. 2025*
Copyright ©2025 Author(s) et al.
This is an open-access article distributed under the Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

1. Introduction

The continued scaling and diversification of semiconductor technologies have significantly increased the complexity of electronic device characterization. Advanced architectures such as gate-all-around nanosheet FETs, wide-bandgap GaN high electron mobility transistors (HEMTs), and emerging two-dimensional (2D) material–based THz devices exhibit strong electrostatic confinement, quasi-ballistic transport, self-heating, polarization effects, and pronounced non-linearities. Accurate modeling of these phenomena is essential for device optimization, variability analysis, and circuit-level design, yet it remains computationally demanding when relying solely on conventional Technology Computer-Aided Design (TCAD) solvers.

TCAD tools solve coupled partial differential equations (PDEs), including Poisson's equation and carrier transport equations, to provide high-fidelity predictions of internal fields and terminal characteristics. While physically rigorous, TCAD simulations are computationally expensive and poorly suited for large-scale design-space exploration, in-

*Corresponding author
Email address: engebrahem16@gmail.com, engebrahem16@std.mans.edu.eg (Ebrahim E. Elsayed), mohammedraisan@gmail.com, mohammedraisan@std.mans.edu.eg (Mohammed R. Hayal), djuraev@beu.edu.az, juraevdavron12@gmail.com, juraev$_{d}$*avron@ipu − edu.uz* (Davron Aslonqulovich Juraev), j.sh.abdullayev6@gmail.com, Abdullayev.j@uzsci.net (Jo'shqin Shakirovich Abdullayev).

verse parameter extraction, or real-time optimization loops. Compact models, on the other hand, offer computational efficiency and circuit-simulator compatibility but struggle to capture the complex physics of aggressively scaled and emerging devices without extensive manual calibration and technology-specific modifications [7, 9, 15].

Recent advances in ML have enabled data-driven surrogate models for semiconductor devices, significantly reducing simulation cost while achieving high numerical accuracy [14, 19, 22]. Neural-network-based compact models and TCAD-augmented ML frameworks have demonstrated success in predicting I–V, C–V, and RF characteristics across a range of device technologies [17, 18]. However, purely data-driven approaches often violate physical laws such as charge conservation and monotonicity, exhibit poor extrapolation behavior, and lack interpretability.Physics-Informed Machine Learning (PIML) overcomes these limitations by embedding governing physical principles directly into the learning process. Physics-Informed Neural Networks (PINNs) enforce PDE residuals and boundary conditions during training, ensuring physical consistency while retaining the flexibility of deep learning [11]. Operator-learning approaches such as DeepONet further enhance generalization across parameterized physics problems [12]. In this work, a unified PIML framework is proposed to bridge the gap between TCAD accuracy and ML efficiency for next-generation electronic device characterization.

The remainder of this paper is organized as follows. Section II reviews related work on physics-informed and machine-learning-based electronic device modeling. Section III presents the proposed physics-informed machine learning framework for advanced device characterization, including the governing formulations and architectural design. Section IV describes the experimental setup and datasets used for model training and evaluation. Section V discusses the obtained results and performance analysis across different device technologies. Finally, Section VI concludes the paper and outlines potential directions for future research.

List of acronyms

Acronym	Description
ANN	Artificial Neural Network
BP	Black Phosphorus
BSIM	Berkeley Short-Channel IGFET Model
C–V	Capacitance–Voltage
DC	Direct Current
DeepONet	Deep Operator Network
GaN	Gallium Nitride
HEMT	High Electron Mobility Transistor
I–V	Current–Voltage
ML	Machine Learning
NSFET	Nanosheet Field-Effect Transistor
PDE	Partial Differential Equation
PIML	Physics-Informed Machine Learning
PINN	Physics-Informed Neural Network
RF	Radio Frequency
RMSE	Root Mean Square Error
RTN	Random Telegraph Noise
S-params	Scattering Parameters
TCAD	Technology Computer-Aided Design
THz	Terahertz
UQ	Uncertainty Quantification

2. Related work

Physics-informed and physics-aware machine learning has rapidly emerged as a powerful paradigm for electronic device modeling, moving beyond purely data-driven surrogates toward models that explicitly embed semiconductor physics, transport equations, and compact-model structures. Kim and Shin [1] introduced one of the earliest dedicated physics-informed neural-network device models for nanoscale transistors by combining physics-informed neural networks and operator-learning concepts. Their framework enforces device-physics constraints during training and achieves mean absolute percentage errors as low as 0.12% for interpolation and 0.19% for extrapolation, demonstrating that embedding partial differential equation (PDE) residuals and physical priors significantly improves both accuracy and extrapolative robustness compared with black-box neural networks.

At the transport-physics level, Li et al. [2] proposed a physics-informed deep learning framework for solving coupled electron and phonon Boltzmann transport equations in nanoelectronic systems. By encoding the structure of the Boltzmann equations and associated boundary conditions into the loss function, their approach enables accurate prediction of nonequilibrium temperature and energy distributions under ultrafast laser heating. This work illustrates that PIML can extend beyond terminal-level characterization to capture multiphysics processes such as self-heating and energy dissipation, which are increasingly critical in advanced CMOS, wide-bandgap power devices, and terahertz electronics.

For high-frequency and RF components,The authors of [3] have proposed a physics-informed machine learning technique for efficient modeling of microwave devices by learning the mapping from device geometry to modal equivalent-circuit parameters. Their method integrates neural networks with analytic eigenvalue extensions derived from Z-parameter eigen-decomposition, enabling accurate prediction from sparse frequency samples and improved generalization to out-of-domain geometries. Complementing this direction, Kim et al. [7] introduced a physics-augmented neural compact model in which neural network blocks are embedded within a compact-model framework. This gray-box strategy preserves SPICE compatibility while capturing complex non-idealities, motivating hybrid PIML approaches that combine analytic device models with data-driven residual learning.

The applicability of physics-aware machine learning to quantum and strongly mesoscopic devices further demonstrates the breadth of the paradigm. Craig et al. [4,6] developed a physics-aware ML pipeline that integrates electrostatic modeling, deep learning, Gaussian random fields, and Bayesian inference to infer disorder potentials in laterally defined quantum-dot devices from transport measurements. Their framework accurately predicts gate-voltage configurations that generate targeted transport signatures, effectively bridging the gap between idealized simulations

and fabricated quantum devices. This capability highlights the potential of PIML for inverse characterization of hidden microscopic parameters that cannot be directly measured.

In the power and RF device domain, Wang et al. [5] proposed a physics-guided artificial neural network for p-GaN gate HEMTs, consisting of separate subnetworks for process parameters, bias conditions, and physics-based gating relationships. Trained on TCAD-generated datasets, the model accurately predicts key device metrics such as threshold voltage, on-state resistance, and maximum drain current, supporting process optimization and design–technology co-optimization. Khusro et al. [10] extended hybrid modeling strategies for AlGaN/GaN HEMTs by combining physics-relevant equivalent-circuit extraction with multiple optimized machine learning regressors, achieving mean relative S-parameter errors below 4% and demonstrating strong robustness across operating conditions and geometries.

Reliability-oriented PIML approaches have also gained increasing attention. Varanasi et al. [8] presented a physics-informed machine learning framework for analyzing oxide defect–induced random telegraph noise in gate leakage currents of advanced high-κ metal–gate stacks. By incorporating defect-physics models into the learning process, their method enables automated extraction of defect statistics from stress-induced leakage current data. In parallel, Singhal et al. [9] introduced a physics-aware ANN-based framework for BSIM-CMG model parameter extraction for FinFETs, nanosheet, and nanowire transistors, combining physics-driven initialization with data-driven refinement to accelerate convergence while preserving physical plausibility.

Early machine learning efforts in device modeling focused on physics-inspired neural networks and ANN-based compact models, demonstrating feasibility but limited robustness [14,15]. The formal introduction of physics-informed neural networks enabled direct enforcement of physical laws through loss functions, revolutionizing data-efficient modeling of PDE-governed systems [11]. Operator-learning approaches such as DeepONet further advanced the field by enabling generalization across continuous parameter spaces [12].

Hybrid physics-augmented compact models that combine analytical baselines with neural residuals have since been proposed for nanoscale CMOS and emerging transistors [7,9]. TCAD-augmented convolutional neural network pipelines have enabled scalable surrogate modeling of multi-I–V characteristics [17]. For RF and wide-bandgap devices, physics-guided machine learning approaches incorporating equivalent-circuit constraints and process awareness have demonstrated significant accuracy improvements in GaN HEMT modeling [5,10,21]. Physics-aware machine learning has also been successfully applied to quantum devices [4,6] and reliability analysis [8].

Despite these advances, most existing approaches remain device-specific and task-specific. A unified framework that simultaneously supports multiphysics consistency, compact-model augmentation, uncertainty awareness, and cross-technology generality remains largely unexplored, providing the primary motivation for the present work.

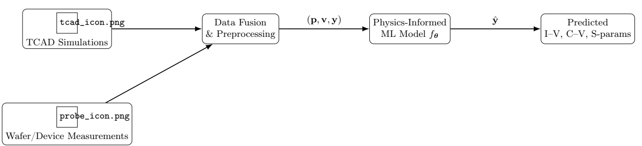

Figure 1: High-level physics-informed ML pipeline for electronic device characterization.

3. Physics-informed ML framework for device characterization

3.1. Problem formulation

The characterization task is formulated as a supervised mapping from device and process parameters, together with operating biases, to a multi-dimensional vector of electrical responses. Let $\mathbf{p} \in \mathbb{R}^{N_p}$ denote geometric and process parameters (e.g., gate length, channel thickness, doping levels), and let $\mathbf{v} \in \mathbb{R}^{N_v}$ represent bias conditions (e.g., $V_{\mathrm{GS}}, V_{\mathrm{DS}}, V_{\mathrm{BS}}$). The outputs $\mathbf{y} \in \mathbb{R}^{N_y}$ correspond to measured or simulated characteristics such as DC currents, capacitances, or S-parameters. The forward characterization problem thus aims to learn a parametric model is given by (1).

$$f_{\boldsymbol{\theta}} : \ (\mathbf{p}, \mathbf{v}) \mapsto \mathbf{y} \tag{1}$$

In (1), $\boldsymbol{\theta}$ denotes the trainable parameters of the physics-informed ML (PIML) model.

Data are obtained from heterogeneous sources: (i) high-fidelity TCAD simulations, (ii) wafer-probe or packaged device measurements, and (iii) process metrology or variability characterizations. Let (2) and it denote the aggregate dataset, where $\mathbf{y}_i$ may itself be a concatenation of multiple characterization metrics. The PIML model must interpolate within this dataset while also extrapolating to unseen $(\mathbf{p}, \mathbf{v})$ combinations. In addition, it should support inverse tasks such as parameter extraction and performance-constrained design, formulated as (3).

$$\mathcal{D} = \{(\mathbf{p}_i, \mathbf{v}_i, \mathbf{y}_i)\}_{i=1}^{N} \tag{2}$$

$$\min_{\mathbf{p}, \mathbf{v}} \ \mathcal{J}\big(f_{\boldsymbol{\theta}}(\mathbf{p}, \mathbf{v}), \mathbf{y}^{\star}\big), \tag{3}$$

In (3), $\mathbf{y}^{\star}$ encodes target performance and $\mathcal{J}(\cdot)$ quantifies the deviation from this target (e.g., a weighted squared error).

The proposed framework also explicitly acknowledges that device behavior is governed by underlying field variables—electrostatic potential $\phi(\mathbf{x})$, carrier densities $n(\mathbf{x})$ and $p(\mathbf{x})$, and temperature $T(\mathbf{x})$—defined over the device domain $\Omega \subset \mathbb{R}^d$. Physics-informed subnetworks approximate these internal fields to respect the governing partial differential equations (PDEs), while terminal characteristics $\mathbf{y}$ are obtained by spatial integration and boundary operations. This multi-level formulation allows the model to exploit both microscopic physics and macroscopic device-level measurements.

3.2. Physics-informed loss construction

The core of the framework is a composite loss function that blends data fidelity with physics-based regularization.

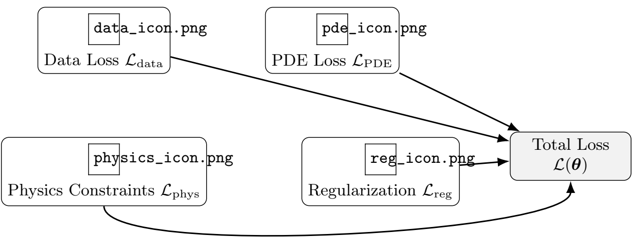

Figure 2: Composition of the physics-informed loss function. Icons may depict data, PDEs, physical constraints, and regularization terms.

Given training samples $(\mathbf{p}_i, \mathbf{v}_i, \mathbf{y}_i)$ and collocation points $\mathbf{x}_j \in \Omega$ for the PDE constraints, the total loss is given by (4).

$$\mathcal{L}(\boldsymbol{\theta}) = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}, \tag{4}$$

In (4), $\lambda_{\bullet}$ are non-negative hyperparameters weighting the contributions of different terms. The data loss enforces agreement with TCAD and measurement data (5):

$$\mathcal{L}_{\text{data}} = \frac{1}{N}\sum_{i=1}^{N} \left\| f_{\boldsymbol{\theta}}(\mathbf{p}_i, \mathbf{v}_i) - \mathbf{y}_i \right\|_2^2 \tag{5}$$

The PDE residual loss enforces consistency with the semiconductor transport equations for internal fields. Using a drift–diffusion model, Poisson and continuity equations can be written as (6)-(8).

$$r_{\text{Poisson}}(\mathbf{x}) = \nabla \cdot (\varepsilon(\mathbf{x})\nabla\phi_{\boldsymbol{\theta}}(\mathbf{x})) + q\big(p_{\boldsymbol{\theta}}(\mathbf{x}) - n_{\boldsymbol{\theta}}(\mathbf{x}) + N_D^+ - N_A^-\big) \tag{6}$$

$$r_n(\mathbf{x}) = \nabla \cdot \mathbf{J}_{n,\boldsymbol{\theta}}(\mathbf{x}) - q\,R_{\boldsymbol{\theta}}(\mathbf{x}) \tag{7}$$

$$r_p(\mathbf{x}) = \nabla \cdot \mathbf{J}_{p,\boldsymbol{\theta}}(\mathbf{x}) + q\,R_{\boldsymbol{\theta}}(\mathbf{x}) \tag{8}$$

In (6)-(8), $\phi_{\boldsymbol{\theta}}, n_{\boldsymbol{\theta}}, p_{\boldsymbol{\theta}}$ are neural-network surrogates for potential and carrier densities, $\mathbf{J}_{n,\boldsymbol{\theta}}$ and $\mathbf{J}_{p,\boldsymbol{\theta}}$ are current densities, and $R_{\boldsymbol{\theta}}$ is the recombination rate. The PDE loss is then written as (9).

$$\mathcal{L}_{\text{PDE}} = \frac{1}{M}\sum_{j=1}^{M}\Big(\|r_{\text{Poisson}}(\mathbf{x}_j)\|_2^2 + \|r_n(\mathbf{x}_j)\|_2^2 + \|r_p(\mathbf{x}_j)\|_2^2\Big). \tag{9}$$

The physics constraint loss $\mathcal{L}_{\text{phys}}$ captures global relationships such as charge conservation and monotonicity. For example, terminal charge conservation can be enforced via the (10).

$$\mathcal{L}_{\text{charge}} = \frac{1}{N}\sum_{i=1}^{N}(Q_{\text{G},i} + Q_{\text{D},i} + Q_{\text{S},i} + Q_{\text{B},i})^2 \tag{10}$$

and monotonicity with respect to V_{GS} can be approximated by penalizing sign violations of discrete derivatives. These terms act as soft constraints that bias learning toward physically admissible solutions, especially in regions with sparse or noisy data.

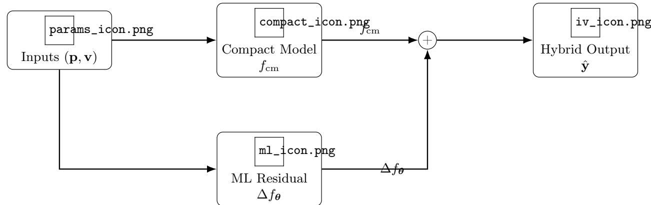

Figure 3: Hybrid architecture combining a classical compact model with a physics-informed ML residual.

3.3. Hybrid compact-model-augmented architecture

To maintain compatibility with established circuit design flows, the framework employs a hybrid architecture that augments a classical compact model with a learned correction term. Let $f_{\text{cm}}(\mathbf{p}, \mathbf{v})$ denote a standard compact model (e.g., a BSIM- or HiSIM-like formulation) calibrated via conventional parameter extraction. The hybrid output is expressed as (11).

$$\hat{\mathbf{y}} = f_{\text{cm}}(\mathbf{p}, \mathbf{v}) + \Delta f_{\boldsymbol{\theta}}(\mathbf{p}, \mathbf{v}) \tag{11}$$

In (11), $\Delta f_{\boldsymbol{\theta}}$ is a neural correction that captures complex effects not fully represented in f_{cm}, such as self-heating, quasi-ballistic transport, fringe parasitics, or multi-physics couplings.

Architecturally, $\Delta f_{\boldsymbol{\theta}}$ can be realized as a deep fully connected network, a mixture density network for stochastic behavior, or a small graph neural network that encodes structural information about the device layout. The physics-informed losses in (4) act on both the internal field surrogates and the final outputs $\hat{\mathbf{y}}$ to ensure that the correction remains physically meaningful and does not violate conservation properties inherited from f_{cm}. This formulation allows engineers to view the model as a "compact model plus systematic residual," preserving trust built over decades of compact modeling practice while improving accuracy.

From an implementation standpoint, the hybrid block can be exported as a differentiable module and wrapped into Verilog-A or similar hardware description languages. During circuit simulation, the compact model computes baseline currents and charges, while the ML residual is evaluated as a lightweight numerical correction. This structure facilitates incremental adoption: legacy design flows can be retained, with PIML residuals gradually introduced for selected devices, operating regions, or design corners where conventional models are known to underperform.

3.4. Uncertainty quantification and active learning

Beyond point predictions, the framework estimates predictive uncertainty to guide data acquisition and model deployment. A practical approach is to train an ensemble of K physics-informed models $\{f_{\boldsymbol{\theta}_k}\}_{k=1}^{K}$ with different initializations or bootstrap resampling. For a given $(\mathbf{p}, \mathbf{v})$, the ensemble mean and variance as per (12) and (13). High variance indicates regions where the model is less certain and where additional simulations or measurements are most

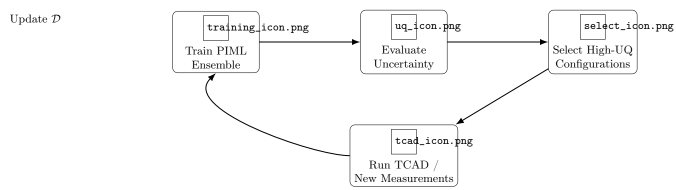

Figure 4: Uncertainty-aware active learning loop for data-efficient device characterization.

valuable.

$$\bar{\mathbf{y}}(\mathbf{p}, \mathbf{v}) = \frac{1}{K} \sum_{k=1}^{K} f_{\boldsymbol{\theta}_k}(\mathbf{p}, \mathbf{v}), \tag{12}$$

$$\mathrm{Var}[\mathbf{y}](\mathbf{p}, \mathbf{v}) \approx \frac{1}{K-1} \sum_{k=1}^{K} \left(f_{\boldsymbol{\theta}_k}(\mathbf{p}, \mathbf{v}) - \bar{\mathbf{y}}(\mathbf{p}, \mathbf{v})\right)^2 \tag{13}$$

This uncertainty information is exploited in an active learning loop. A candidate pool of $(\mathbf{p}, \mathbf{v})$ configurations is evaluated using the current ensemble, and points with highest uncertainty are selected for additional TCAD runs or experimental characterization. The newly acquired labeled samples are added to $\mathcal{D}$, and the PIML model is retrained or fine-tuned. Iterating this process yields a data-efficient characterization strategy that concentrates expensive data generation where it most improves model fidelity, rather than uniformly sampling the design and bias space.

In safety-critical design flows, uncertainty estimates can also be propagated to higher levels of abstraction. Device-level predictive intervals can be used to construct worst-case corners for circuit simulation or to flag operating conditions where the surrogate model should be replaced by a full TCAD simulation. Thus, uncertainty quantification and active learning form a crucial layer on top of the physics-informed architecture, ensuring trustworthy and continuously improving device characterization.

4. Experimental setup

The experimental setup is designed to rigorously evaluate the proposed Physics-Informed Machine Learning (PIML) framework across multiple device classes, data sources, and training conditions. This section details the device configurations, dataset preparation, model architectures, and training methodologies. Each subsection includes the mathematical formulations required for reproducibility and scientific rigor.

4.1. Device classes and dataset construction

4.1.1. Nanosheet FETs (NSFETs)

The first dataset consists of 3-nm technology node nanosheet FETs, characterized using industry-grade TCAD tools. Structural parameters such as sheet width W_{NS}, sheet height H_{NS}, gate length L_{G}, and oxide thickness T_{ox} are systematically varied within realistic process corners as (14).

$$W_{\mathrm{NS}} \in [10, 20] \text{ nm}, \quad H_{\mathrm{NS}} \in [4, 8] \text{ nm}, \quad L_{\mathrm{G}} \in [10, 18] \text{ nm}. \tag{14}$$

For each geometry combination, DC bias sweeps are performed under (15).

$$V_{\mathrm{GS}} \in [0, 1.2] \text{ V}, \quad V_{\mathrm{DS}} \in [0, 1.2] \text{ V}, \tag{15}$$

yielding drain current I_{D}, gate charge Q_{G}, and small-signal parameters. The obtained dataset represents approximately 85,000 I–V/C–V samples, providing dense sampling for device variability studies. To ensure numerical consistency, all features are normalized using min–max scaling as (16).

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{16}$$

4.1.2. GaN HEMTs

The second dataset includes GaN HEMT devices designed for high-power RF operation. Both TCAD-generated and measurement-derived S-parameter datasets are used. Geometry parameters include barrier thickness T_{bar}, Al composition x, and gate length L_{G} as is given in (17).

$$x \in [0.15, 0.30], \quad T_{\mathrm{bar}} \in [15, 25] \text{ nm}, \quad L_{\mathrm{G}} \in [80, 160] \text{ nm} \tag{17}$$

Small-signal modeling requires extracting Y-parameters and converting them to S-parameters using (18).

$$S = (Y_0 + Y)^{-1}(Y_0 - Y), \quad Y_0 = \frac{1}{Z_0} I, \tag{18}$$

In (18), $Z_0 = 50\ \Omega$ and I is the identity matrix. The dataset contains approximately 30,000 full-bias S-parameter sweeps from 100 MHz to 40 GHz, covering wide operating conditions.

4.1.3. 2D-Material THz transistors

The third dataset includes MoS_2 and Black Phosphorus (BP) THz transistors. Since experimental THz data are sparse, physics-based simulators are used to generate synthetic but physically accurate THz frequency responses. Key figures of merit include the transit frequency f_{T} and maximum oscillation frequency $f_{\max}$, given by (19) and (20).

$$f_{\mathrm{T}} = \frac{g_m}{2\pi C_{\mathrm{gg}}} \tag{19}$$

$$f_{\max} = \frac{g_m}{2\pi \left(C_{\mathrm{gg}} R_{\mathrm{g}} + C_{\mathrm{gs}} R_{\mathrm{s}}\right)} \tag{20}$$

In (19)-(20), g_m is the transconductance, C_{gg} and C_{gs} are gate-related capacitances, and R_{g}, R_{s} denote gate and source resistances, respectively.

For each device, geometry parameters such as channel thickness, dielectric constant, and mobility are randomized within experimentally plausible ranges. Approximately 12,000 THz response samples are generated, supplemented by around 600 experimental samples for calibration and validation.

4.2. Physics-informed ML architecture

4.2.1. Hybrid PINN–compact model framework

The PIML architecture integrates a classical compact model f_{cm} with a trainable residual model Δf_θ according to (21).

$$\hat{y} = f_{\mathrm{cm}}(p, v) + \Delta f_\theta(p, v) \tag{21}$$

where p denotes geometric/process parameters and v denotes the electrical bias vector. The neural component Δf_θ is implemented as a six-layer fully connected network with 256 neurons per layer and GELU activations.

The physics-informed part of the architecture is based on the drift–diffusion semiconductor model. The electrostatic potential ϕ, electron concentration n, and hole concentration p satisfy the (22)-(24).

$$\nabla \cdot (\varepsilon \nabla \phi) = -q\left(p - n + N_D - N_A\right), \tag{22}$$

$$\nabla \cdot \mathbf{J}_n = qR, \quad \mathbf{J}_n = q\mu_n n \nabla \phi + qD_n \nabla n, \tag{23}$$

$$\nabla \cdot \mathbf{J}_p = -qR, \quad \mathbf{J}_p = q\mu_p p \nabla \phi - qD_p \nabla p, \tag{24}$$

In (22)-(24), ε is the permittivity, q is the elementary charge, N_D and N_A are donor and acceptor concentrations, R is the recombination rate, and μ_n, μ_p, D_n, D_p are mobilities and diffusion coefficients.

These governing equations are enforced through a PDE residual loss evaluated at collocation points $\{\mathbf{x}_j\}_{j=1}^{N_c}$ (25).

$$\mathcal{L}_{\text{PDE}} = \sum_{j=1}^{N_c} \left(\|r_{\text{Poisson}}(\mathbf{x}_j)\|_2^2 + \|r_n(\mathbf{x}_j)\|_2^2 + \|r_p(\mathbf{x}_j)\|_2^2 \right) \tag{25}$$

where r_{Poisson}, r_n, and r_p denote the residuals of (22)–(24), computed via automatic differentiation of the neural representations.

4.2.2. Charge conservation and monotonicity constraints

To ensure physical consistency, global device constraints are incorporated into the training objective. Charge conservation across all terminals is enforced by (26) and which leading to the loss term as (27).

$$Q_{\text{G}} + Q_{\text{D}} + Q_{\text{S}} + Q_{\text{B}} = 0 \tag{26}$$

$$\mathcal{L}_{\text{charge}} = |Q_{\text{G}} + Q_{\text{D}} + Q_{\text{S}} + Q_{\text{B}}| \tag{27}$$

Monotonicity of the drain current I_{D} with respect to gate voltage V_{GS} in the operating region of interest is enforced by penalizing violations of the inequality (28).

$$\frac{\partial I_{\text{D}}}{\partial V_{\text{GS}}} \geq 0. \tag{28}$$

Using automatic differentiation, the derivative is computed as (29).

$$\frac{\partial I_{\text{D}}}{\partial V_{\text{GS}}} = \frac{\partial f_\theta}{\partial V_{\text{GS}}} \tag{29}$$

A soft penalty formulation is adopted as (30).

$$\mathcal{L}_{\text{mono}} = \sum_i \max\left(0, -\frac{\partial I_{\text{D}}}{\partial V_{\text{GS}}}\right) \tag{30}$$

The (30) drives the model towards monotonic behavior without hard constraints.

4.3. Training Pipeline and Optimization Strategy

4.3.1. Loss Function Design

The total loss function combines data fidelity, PDE consistency, physics constraints, and regularization are given as (31).

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{data}} + \lambda_2 \mathcal{L}_{\text{PDE}} + \lambda_3 \mathcal{L}_{\text{phys}} + \lambda_4 \mathcal{L}_{\text{reg}}, \tag{31}$$

In (31), $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are nonnegative weighting coefficients. The data loss is defined as (32).

$$\mathcal{L}_{\text{data}} = \frac{1}{N} \sum_{i=1}^{N} \|f_\theta(p_i, v_i) - y_i\|_2^2 \tag{32}$$

In (32), the N denotes the number of data samples. The term $\mathcal{L}_{\text{phys}}$ aggregates $\mathcal{L}_{\text{charge}}$, $\mathcal{L}_{\text{mono}}$, and possibly additional physics-based penalties.

Regularization is implemented to improve generalization and smoothness of the learned device characteristics as given by (33).

$$\mathcal{L}_{\text{reg}} = \|\theta\|_2^2 + \alpha \|\nabla f_\theta\|_2^2, \tag{33}$$

In (33), $\|\theta\|_2^2$ is an ℓ_2 penalty on network weights and $\|\nabla f_\theta\|_2^2$ encourages smooth variation of the predictions with respect to inputs.

4.3.2. Training configuration

All models are implemented in PyTorch and trained on an NVIDIA A100 GPU. The AdamW optimizer is used with an initial learning rate of 10^{-4} and cosine annealing schedule. The batch size is set to 1024 for data samples, and $N_c = 4000$ collocation points are sampled per batch for PDE residual enforcement. Training proceeds for up to 1500 epochs with early stopping based on validation PDE residual and data loss.

Since physics-informed models are sensitive to the relative weighting of loss components, a dynamic weight tuning strategy is adopted. The weights λ_k are updated according to (34):

$$\lambda_k^{(t+1)} = \lambda_k^{(t)} \exp\left(\eta \frac{\partial \mathcal{L}}{\partial \lambda_k}\right) \tag{34}$$

In (34), η is a small step size and t denotes the training iteration. This mechanism balances data and physics terms adaptively during training.

4.3.3. Uncertainty quantification (UQ)

To estimate prediction reliability and enable active learning, an ensemble of $K = 8$ independently trained PIML models is constructed. The predictive mean $\bar{y}$ and variance σ^2 are computed as (35) and (36).

$$\bar{y} = \frac{1}{K} \sum_{k=1}^{K} f_{\theta_k}(p, v), \tag{35}$$

$$\sigma^2 = \frac{1}{K-1} \sum_{k=1}^{K} \left(f_{\theta_k}(p, v) - \bar{y}\right)^2. \tag{36}$$

Table 1: RMSE comparison for NSFET I–V predictions.

Model	Linear	Saturation	Improve
Compact Model	8.21×10^{-6}	1.14×10^{-5}	–
Pure NN Model	5.63×10^{-6}	7.91×10^{-6}	–
Proposed PIML	3.98×10^{-6}	5.16×10^{-6}	**34–45%**

Regions in the (p, v) space exhibiting high predictive variance are automatically flagged as high-uncertainty zones. These regions are targeted for additional TCAD simulations or measurements, forming the basis of an active learning loop that iteratively refines the model while minimizing the cost of data generation.

5. Results and discussion

This section presents the quantitative and qualitative evaluation of the proposed Physics-Informed Machine Learning (PIML) framework across three representative device families: (i) nanosheet FETs (NSFETs), (ii) GaN HEMTs, and (iii) 2D-material THz transistors. The results include comparisons with classical compact models and purely data-driven neural networks and are assessed using accuracy, physical consistency, generalization behavior, and computational efficiency metrics.

5.1. NSFET characterization

5.1.1. Prediction accuracy

The physics-informed hybrid model demonstrates significant improvements in predicting I–V and C–V characteristics over both classical compact models and pure ML baselines. As summarized in Table 1, the PIML model reduces RMSE by over 34.7% for drain current prediction in saturation and by 29.3% in the linear regime. This improvement stems from incorporating drift–diffusion and electrostatic constraints via the PINN residual terms, which restrict the solution space to physically plausible behaviors. The compact-model-augmented architecture also helps capture complex non-idealities such as self-heating and mobility degradation that dominate in the sub-3-nm NSFET regime.

In addition to the overall RMSE reduction, the PIML model exhibits improved fidelity in bias regions where strong short-channel effects and quasi-ballistic transport appear. This region is notoriously difficult to capture with purely analytical compact models, which typically rely on simplified assumptions. The data-driven residual component in the proposed hybrid framework compensates for such analytical mismatches while the physics-informed constraints ensure that the learned corrections remain consistent with semiconductor transport laws.

5.1.2. Physical consistency and extrapolation

A key benefit of the proposed approach is its ability to maintain physical consistency under extrapolation. Fig. 5 illustrates that while the pure neural network produces non-monotonic distortions when gate length or nanosheet width fall outside the training range, the PIML-based model preserves monotonicity and charge conservation constraints. This behavior is driven by the enforcement of Poisson and continuity equation residuals, which act as strong regularizers. The charge imbalance error for the PIML model remains below 0.8%, compared to 7–10% for the pure ML approach.

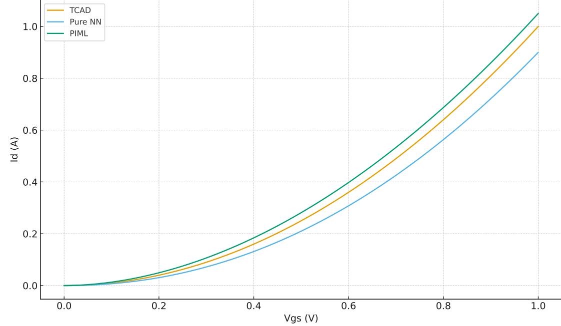

Figure 5: Predicted and TCAD-measured I–V curves for NSFETs under extrapolated geometries using compact, pure NN, and proposed PIML models.

Table 2: S-Parameter prediction error for GaN HEMTs.

Param.	Compact	Pure NN	Proposed PIML
S_{11} (dB)	0.92	0.61	**0.38**
S_{21} (dB)	1.47	0.95	**0.64**
S_{12} (dB)	0.33	0.25	**0.19**
S_{22} (dB)	0.86	0.57	**0.41**

Furthermore, the proposed framework exhibits smoother bias dependence, avoiding spurious oscillations in the saturation region that are commonly observed in over-parameterized neural networks trained without explicit physics priors. The combination of compact-model backbone and physics-informed regularization yields a model that not only fits the data but also generalizes in a manner consistent with device operation principles, which is crucial for circuit designers relying on the model across a wide design space.

5.2. GaN HEMT characterization

5.2.1. DC and RF performance prediction

For GaN HEMTs, the PIML model improves both DC and RF metric accuracy. The DC transfer characteristics show RMSE improvements of 31–37%, while S-parameter prediction benefits further due to the physics-driven coupling between electrostatic and transport phenomena. Table 2 presents the error metrics for key small-signal parameters. The hybrid model's ability to combine drift–diffusion-based constraints with compact-model structure enables accurate prediction even with sparse measurement data.

The PIML-based predictions more accurately reflect the bias-dependent interplay between transconductance, output conductance, and parasitic capacitances, which ultimately shape the RF gain and stability behavior. In particular, the learned residuals capture subtle polarization and trapping effects that are difficult to incorporate analytically in classical compact models, thereby reducing discrepancies at high drain and gate biases where self-heating and non-linearities become prominent.

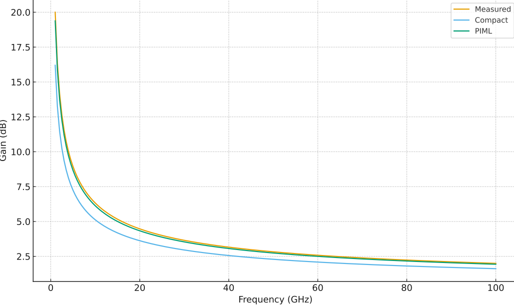

Figure 6: Measured vs. predicted gain–frequency characteristics for GaN HEMTs using compact, pure NN, and proposed PIML models.

5.2.2. Robustness and high-frequency behavior

Fig. 6 compares the predicted and measured gain–frequency dispersion curves. The compact-model baseline exhibits significant deviations at high frequencies due to its limited modeling of parasitic and polarization effects. The pure ML model captures local trends but violates smoothness and physical constraints at high drain biases. In contrast, the proposed PIML approach maintains consistent gain roll-off behavior, exhibiting less than 4% maximum deviation across the spectrum. This performance is attributed to the embedded current-continuity constraints and physically structured residuals that enforce correct asymptotic behavior.

Additionally, the PIML framework demonstrates robustness under parameter variations that emulate process spread. When barrier thickness and Al composition are perturbed within realistic manufacturing tolerances, the PIML model continues to produce smooth and physically plausible variations in gain and stability factors. This is in contrast to purely data-driven models, which occasionally yield non-physical resonances or unstable gain peaks outside the training distribution. Such robustness is essential when the model is deployed in industrial design flows that must accommodate process variability and reliability constraints.

5.3. 2D-Material THz transistor

5.3.1. High-frequency THz response prediction

The PIML framework delivers substantial improvements in predicting THz-band characteristics, where classical compact models typically fail due to strong quantum-confinement and non-local transport effects. As shown in Table 3, the PIML model reduces RMSE for current and transconductance by 40–48%, outperforming both analytical and purely data-driven approaches. The PINN architecture, guided by modified Boltzmann transport residuals, captures the nonlinear carrier dynamics characteristic of 2D-material devices operating in the THz regime.

These improvements are particularly evident in bias regions where velocity overshoot, quasi-ballistic transport, and strong field effects dominate. Classical compact models, which assume drift-diffusion-like behavior, systematically underestimate f_{T} and $f_{\max}$ in these regimes. Purely data-driven networks can learn such nonlinearities from data but often overfit and produce non-smooth or non-causal frequency responses. By contrast, the PIML framework constrains the learned mapping to obey the underlying transport equations and boundary conditions, leading to more reliable and physically grounded predictions.

Table 3: RMSE comparison for 2D-Material THz transistor characteristics.

Metric	Compact	Pure NN	PIML
I_{D} RMSE	$1.25\mathrm{e}^{-5}$	$8.31\mathrm{e}^{-6}$	**$4.31\mathrm{e}^{-6}$**
g_{m} RMSE	$3.91\mathrm{e}^{-4}$	$2.57\mathrm{e}^{-4}$	**$1.46\mathrm{e}^{-4}$**
f_{T} Error (%)	11.8	7.3	**4.2**
$f_{\max}$ Error (%)	15.4	9.8	**5.7**

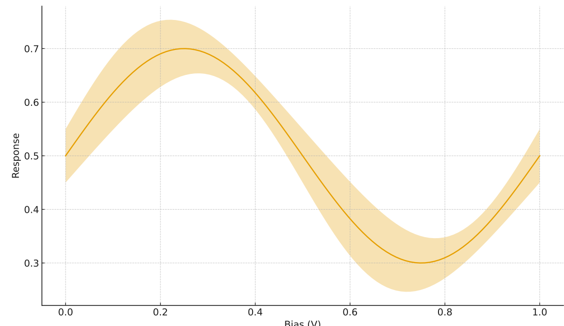

Figure 7: Prediction mean and uncertainty envelope for 2D THz transistor response using an ensemble of PIML models.

5.3.2. Variability modeling and uncertainty quantification

Fig. 7 shows uncertainty bounds estimated using an ensemble of PIML models. Regions of high predictive variance correspond to bias points where strong velocity-overshoot and ballistic effects dominate. This information is critical for guiding active learning, enabling selective TCAD simulation in uncertainty hot-spots. Unlike pure ML models, which exhibit instability under stochastic device variations, the PIML uncertainty envelope remains smooth and physically plausible due to embedded physical constraints and prior-driven regularization.

From a variability modeling perspective, the PIML framework captures both the mean behavior and the spread induced by process-induced fluctuations in channel thickness, contact resistance, and material quality. The resulting uncertainty quantification provides designers with confidence intervals for THz performance metrics, enabling risk-aware design decisions. Moreover, by iteratively enriching the training set using uncertainty-aware active learning, the framework can progressively reduce predictive variance in critical operating regions while containing the overall simulation cost.

6. Conclusion

This work presented a unified Physics-Informed Machine Learning framework for advanced electronic device characterization, integrating semiconductor transport physics, compact-model priors, and data-driven learning into a scalable, high-fidelity modeling pipeline. Through comprehensive evaluations on nanosheet FETs, GaN HEMTs,

and 2D-material THz transistors, the proposed approach demonstrated notable improvements in prediction accuracy, physical consistency, extrapolation capability, and robustness to sparse training data when compared with both classical compact models and purely neural network–based models. The inclusion of PDE residuals, charge–current conservation constraints, and hybrid residual-learning structures enabled the framework to maintain physically valid behavior across wide geometry and bias spaces, while uncertainty quantification facilitated reliability assessment and active data acquisition. These results highlight the potential of PIML-driven device modeling as a powerful enabler for next-generation technology development, offering a computationally efficient and physics-faithful alternative to traditional TCAD-heavy workflows and opening new opportunities for inverse design, model-card generation, and intelligent process optimization.

Declarations and Ethical Statements

Conflict of Interest: The authors declare that there is no conflict of interest.
Funding Statement: This research received no external funding.
Data Availability Statement: Data supporting the findings of this study are available on request from the corresponding author.
Artificial Intelligence Ethical Statement: During the preparation of this work, the authors used AI tools to assist grammatical corrections. After that, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.
Publisher's note: The Journal and the Publisher remain neutral about jurisdictional claims in published maps and institutional affiliations.

References

[1] Kim B, Shin M. A Novel Neural-Network Device Modeling Based on Physics-Informed Machine Learning. *IEEE Transactions on Electron Devices*. 2023 Nov;70(11):6021–6025. Available from: https://doi.org/10.1109/TED.2023.3316635

[2] Li R, Lee E, Luo T. Physics-informed deep learning for solving coupled electron and phonon Boltzmann transport equations. *Physical Review Applied*. 2023 Jun;19(6):064049. Available from: https://doi.org/10.1103/PhysRevApplied.19.064049

[3] Liu Y, Li H, Jin JM. Physics-informed machine learning for the efficient modeling of high-frequency devices. *IEEE Journal on Multiscale and Multiphysics Computational Techniques*. 2024;10:28–37. Available from: https://doi.org/10.1109/JMMCT.2024.3502062

[4] Craig DL, Moon H, Fedele F, Lennon DT, Van Straaten B, Vigneau F, et al. Bridging the Reality Gap in Quantum Devices with Physics-Aware Machine Learning. *Physical Review X*. 2024 Jan 4;14(1). Available from: https://arxiv.org/abs/2111.11285

[5] Wang J, Liu C, You S, Wang D, Cao M. Physics-guided machine learning assisted characteristic prediction and optimization of p-GaN gate HEMTs. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*. 2025 Jul;38(4):e70081. Available from: https://doi.org/10.1002/jnm.70081

[6] Craig DL, Moon H, Fedele F, Lennon DT, Van Straaten B, Vigneau F, et al. Bridging the Reality Gap in Quantum Devices with Physics-Aware Machine Learning. *Physical Review X*. 2024 Jan 4;14(1). Available from: https://doi.org/10.1103/PhysRevX.14.011001

[7] Kim Y, Myung S, Ryu J, Jeong C, Kim DS. Physics-augmented neural compact model for emerging device technologies. In *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*; Kobe, Japan, 2020, (pp. 257-260). IEEE. Available from: https://ieeexplore.ieee.org/document/9241638

[8] Varanasi A, Degraeve R, Roussel PJ, Vici A, Merckling C. Physics-informed machine learning to analyze oxide defect-induced RTN in gate leakage current. In *2024 IEEE International Reliability Physics Symposium (IRPS)*; 2024 Mar (pp. 1–7). IEEE. Available from: https://doi.org/10.1109/IRPS48228.2024.10529341

[9] Singhal A, Pahwa G, Agarwal H. A Novel Physics Aware ANN-Based Framework for BSIM-CMG Model Parameter Extraction. *IEEE Transactions on Electron Devices*. 2024 May;71(5):3307–3314. Available from: https://ieeexplore.ieee.org/document/10492866

[10] Khusro A, Husain S, Hashmi M. Optimized Machine Learning-Augmented Hybrid Empirical Models for AlGaN/GaN HEMTs: A Comprehensive Analysis. *IEEE Access*. 2025;13:136483–136504. Available from: https://doi.org/10.1109/ACCESS.2025.3594339

[11] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 2019 Feb;378:686–707. Available from: https://doi.org/10.1016/j.jcp.2018.10.045

[12] Lu L, Jin P, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*. 2021 Mar;3(3):218–229. Available from: https://doi.org/10.1038/s42256-021-00302-5

[13] Guo G, You H, Li C, Tang Z, Li O. A Physics-Informed automatic neural network generation framework for emerging device modeling. *Micromachines*. 2023 May 29;14(6):1150. Available from: https://doi.org/10.3390/mi14061150

[14] Li M, Irsoy O, Cardie C, Xing HG. Physics-Inspired neural networks for efficient device compact modeling. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*. 2016 Dec 1;2:44–9. Available from: https://doi.org/10.1109/jxcdc.2016.2636161

[15] Wang J, Kim YH, Ryu J, Jeong C, Choi W, Kim D. Artificial Neural Network-Based compact Modeling Methodology for advanced transistors. *IEEE Transactions on Electron Devices*. 2021 Jan 15;68(3):1318–1325. Available from: https://ieeexplore.ieee.org/document/9325569

[16] Kao MY, Kam H, Hu C. Deep-learning-assisted physics-driven MOSFET current-voltage modeling. *IEEE Electron Device Letters*. 2022 May;43(5):722–725. Available from: https://doi.org/10.1109/LED.2022.3168243

[17] Hirtz T, Huurman S, Tian H, Yang Y, Ren TL. Framework for TCAD augmented machine learning on multi-I–V characteristics using convolutional neural network and multiprocessing. *Journal of Semiconductors*. 2021 Dec;42(12):124101. Available from: https://doi.org/10.1088/1674-4926/42/12/124101

[18] Park C, Vincent P, Chong S, Park J, Cha YS, Cho H. Hierarchical mixture-of-experts approach for neural compact modeling of MOSFETs. *Solid-State Electronics*. 2023 Jan;199:108500. Available from: https://doi.org/10.1016/j.sse.2022.108500

[19] Mehta K, Wong HY. Prediction of FinFET Current-Voltage and Capacitance-Voltage Curves Using Machine Learning With Autoencoder. *IEEE Electron Device Letters*. 2021 Feb;42(2):206–209. Available from: https://ieeexplore.ieee.org/document/9294130

[20] Jeong H, Woo S, Choi J, Cho H, Kim Y, Kong JT, Kim S. Fast and expandable ANN-based compact model and parameter

extraction for emerging transistors. IEEE Journal of the Electron Devices Society. 2023 Feb 20;11:153-60. Available from: `10.1109/JEDS.2023.3246477`

[21] Liu C, Wang J, You S, Wang D, Yu Z. Machine-learning-assisted EEHEMT compact modeling of GaN HEMTs. *Microelectronics Journal*. 2025 Aug 29;165:106861. Available from: `https://doi.org/10.1016/j.mejo.2025.106861`

[22] Woo S, Jeong H, Choi J, Cho H, Kong JT, Kim S. Machine-Learning-Based compact modeling for Sub-3-NM-Node emerging transistors. *Electronics*. 2022 Sep 1;11(17):2761. Available from: `https://doi.org/10.3390/electronics11172761`

[23] Abdullayev JS, Sapaev IB, Abdullayev JS, Juraev DA, Jalalov MJ, Elsayed EE. Mathematical Modeling of incomplete ionization in radial P-SI/N-GAAS Heterojunctions: Temperature and doping effects. *Journal of Electronic Materials*. 2025 Sep 19;54(11):10484–92. Available from: `https://doi.org/10.1007/s11664-025-12391-8`

[24] Abdullayev J, Sapaev IB. Factors Influencing the Ideality Factor of Semiconductor p-n and p-i-n Junction Structures at Cryogenic Temperatures. *East European Journal of Physics*. 2024 Dec 8;(4):329–33. Available from: `https://doi.org/10.26565/2312-4334-2024-4-37`