A Bottom-Up Machine-Learning Approach for Efficient Device Simulation

Xiaoxin Xie, Yuchen Wang, Zili Tang, Yijiao Wang, *Member, IEEE*, Xing Zhang, and Fei Liu, *Member, IEEE*

***Abstract*—The combination of machine-learning (ML) and electronic structure computation has proven effective in studying various properties of molecules and crystals at the atomistic level. However, challenges arise when these molecules or crystals are contacted with external electrodes, complicating the description of quantum transport properties using existing methods. In this study, we propose an attention-based heterogeneous graph neural network to characterize the global field and dynamic features of open systems. Our approach aims to accelerate or bypass the resource-intensive self-consistent iterations of solving Schrödinger and Poisson equations within nonequilibrium Green's function (NEGF) formalism from the bottom-up, significantly improving the efficiency of quantum transport calculations. Representing the device with a heterogeneous graph largely retains its intrinsic physical characteristics, while the global graph attention network (GAT) effectively captures the propagation of nonlocal physical information, addressing prediction accuracy challenges due to device scaling. The global field heterogeneous graph neural network (GFGNN) demonstrates high accuracy, significant acceleration, and potential transferability at different channel lengths in simulations of p-n junctions (two-terminal with significant tunneling effect) and MOSFETs (three-terminal).**

***Index Terms*—Graph attention network (GAT), machine learning (ML), open system, quantum transport.**

I. Introduction

As TRANSISTOR sizes approach physical limits [1], [2], new materials, structures, and principles are introduced for devices [3], [4], [5], [6]. There is a growing need for advanced quantum and atomic-level simulation tools to provide predictive insights and theoretical guidance. The nonequilibrium Green's function (NEGF) method is an important tool for calculating the quantum transport properties of nanoscale devices [7], [8]. However, the self-consistent iteration of the Schrödinger and Poisson equations within the NEGF framework demands significant computational resources, making the acceleration of NEGF methods essential for efficient device simulation.

Artificial intelligence (AI) is progressively being incorporated into scientific discovery, enhancing and speeding up research processes [9], [10], such as for device compact modeling [11], [12], [13]. Introducing a data-driven research paradigm [14], [15] into quantum transport computing is therefore logical. This approach leverages machine learning (ML) to extract scientific insights from vast amounts of simulation data, thus guiding and accelerating device simulations. The research on ML for the NEGF framework is still in its nascent stage. Among the existing studies, several initial works focus on applying ML techniques to accelerate NEGF calculations [16], [17], [18], [19], [20]. A bottom-up approach that combines ML and NEGF is still necessary, incorporating detailed grid or atom information for quantum transport device simulations.

In recent developments, there is a growing interest in employing advanced ML methods in atomistic material property calculations. For instance, ML has been employed in the mapping of properties [21], [22], the construction of interatomic potentials [23], [24], and the learning of electronic density functional/Hamiltonian [25], [26], [27]. Electronic structure computation focuses on equilibrium properties of confined/periodic systems, while NEGF calculations address transport properties of nonequilibrium open systems. The generalization of ML methodologies, from electronic structure computation of a confined/periodic system to quantum transport calculation of an open system, requires new methodologies in varying contexts, crucial for device research. The advancements of ML-driven NEGF simulations from the bottom-up promise to enhance the applicability, speed, and accuracy of semiconductor device simulations.

In this work, we propose an attention-based global field heterogeneous graph neural network (GFGNN) to characterize field effects and dynamics in open systems using detailed grid or atom information. The GFGNN is trained on a heterogeneous graph derived from the device structure, containing information on electrodes/boundaries, atom types, doping, potentials, and other relevant data. The training objective is to obtain the device's self-consistent potential distribution. We tested our method on two- and three-terminal systems

Received 14 December 2024; revised 10 January 2025; accepted 15 January 2025. This work was supported in part by the National Natural Science Foundation of China under Grant T2293702, Grant T2293703, and Grant T2293700; and in part by the 111 Project under Grant B18001. The review of this article was arranged by Editor J. Guo. *(Corresponding author: Fei Liu.)*

Xiaoxin Xie, Yuchen Wang, Zili Tang, Xing Zhang, and Fei Liu are with the School of Integrated Circuit and Beijing Advanced Innovation Center for Integrated Circuits, Peking University, Beijing 100871, China (e-mail: feiliu@pku.edu.cn).

Yijiao Wang is with the School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100191, China, and also with the National Key Laboratory of Spintronics, Institute of International Innovation, Beihang University, Hangzhou 311115, China.

Digital Object Identifier 10.1109/TED.2025.3533465

using graphene-nanoribbon (GNR)-based p-n junctions and monolayer MoS_2-based MOSFETs. The GFGNN showed an excellent ability to predict different voltage points for unknown device structures. The acceleration of the NEGF calculation process ranges from 174.63% to 418.18% (with self-consistent iteration) and from 971.43% to 1271.43% (without self-consistent iteration), while maintaining accurate transport calculations, demonstrating the GFGNN's strong generalization ability and grasp of intrinsic physical principles.

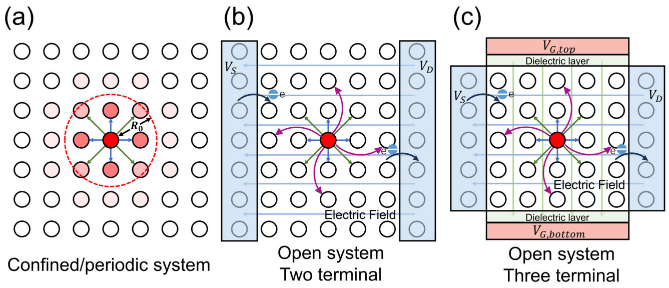

Fig. 1. (a) Confined/periodic system: the influence of atoms within a radius R_0 around the central atom is considered. (b) Open system with two terminals: the source and drain electrodes generate a global electric field and facilitate electron exchange. (c) Open system with three terminals: the source, drain, and gate electrodes are present. The gate introduces an additional electric field, but does not facilitate electron exchange.

II. Confined/Periodic System and Open System

In the study of electronic properties of 1-D, 2-D, and 3-D materials using first-principles calculations or the tight-binding (TB) model, as well as 0-D quantum dots and molecules, periodic or hard boundary conditions must be applied. Conversely, devices function as open systems in which materials are contacted with external electrodes. The interactions with external electrodes, phonons, or photons are described by self-energies within the NEGF formalism. In transport calculations of an open system using NEGF, the rest of the device, excluding the contacts, is described by a Hamiltonian constructed using density functional theory (DFT), TB, k·p model, and effective mass (EM) methods. This approach aligns with those used in electronic structure calculations [28], [29], [30], [31].

The DFT-based device Hamiltonian under a bias voltage must be calculated self-consistently, a notably time-consuming process. In TB- and k·p-based device calculations, the device Hamiltonian is divided into H_0, which remains unchanged, and the potential across the device, which must also be calculated self-consistently and is similarly time-consuming. NEGF calculations necessitate a complex self-consistent solution, with each cycle involving a recursive computation of self-energy and large-scale matrix inversion. This iterative process is the primary source of complexity in transport calculations. In contrast, the subsequent nonself-consistent steps do not require significant computational resources. Therefore, predicting the self-consistent Hamiltonian or potential of the device is a crucial research focus in the field of ML for device transport computation. Extensive works have been explored to obtain the periodic and hard boundary Hamiltonian for electronic structure calculations [25]. While little work has been done to obtain the Hamiltonian or potential for open systems, this is our focus.

We next examine the distinctions and connections between confined/periodic and open systems, emphasizing their application in neural network algorithms. When calculating the electronic structure of confined or periodic systems using localized orbital-based DFT [25] or TB, the influence of each atom on the central atom can be approximated by considering only the atoms within a surrounding radius R_0 [see Fig. 1(a)] [21], [25]. Regardless of the system's size, there are a finite number of central atom types and corresponding local environments. By iterating through all these types during training, we can decompose a large system into numerous subproblems, each characterized by a central atom and its local environment. Summarizing the contributions of each subproblem enables us to predict certain properties of the entire system [25].

In open systems, as illustrated in Fig. 1(b) and (c), the electrodes can exchange electrons with the atoms in the channel and exert a global electric field on these atoms. If we were to adopt the same approach used for confined/periodic boundary conditions, two main challenges would arise.

1) *Boundary Condition Representation:* There is a lack of methods to effectively characterize the different boundary conditions imposed by the contact electrodes. Specifically, it is challenging to describe the semi-infinite source-drain electrodes, which involve both electron exchange and electric field action, and the gate electrodes, which involve only electric field action.
2) *Limitations of Local Approximation:* If each central atom only considers the local environment within a surrounding radius R_0, only the atoms near the boundary will directly experience the influence of the electrodes. However, all atoms are affected by the electrical field. These boundary atoms, once affected by a specific potential, can transfer the influence of the electrodes layer by layer to the inner atoms, creating a complex global dynamic process. This interdependence introduces a contradiction between the complexity and generalization of the network.

Thus, in an open system, it is imperative to develop novel methodologies to address these challenges. We will elaborate on our approach from the perspectives of characterization techniques and network structure.

III. Heterogeneous Graph of Two- and Three-Terminal Devices

In the study of neural networks, it is crucial to adequately characterize the input data. In previous work on atomistic electronic structure prediction, graph structures have shown powerful representational capabilities, and graph-based neural networks have achieved excellent performance [21], [32]. As we know, open system transport simulations are a natural extension of confined/periodic system electronic structure simulations. Therefore, in open systems, our solution is to characterize the device using a heterogeneous graph that can be fully learned in the GFGNN while fully characterizing the intrinsic features of the device [33], [34]. This is shown in Fig. 2. We first define each atom in the device as a device node (represented in blue and green in the figure). Each device node contains three features: atom type (Z), doping (D),

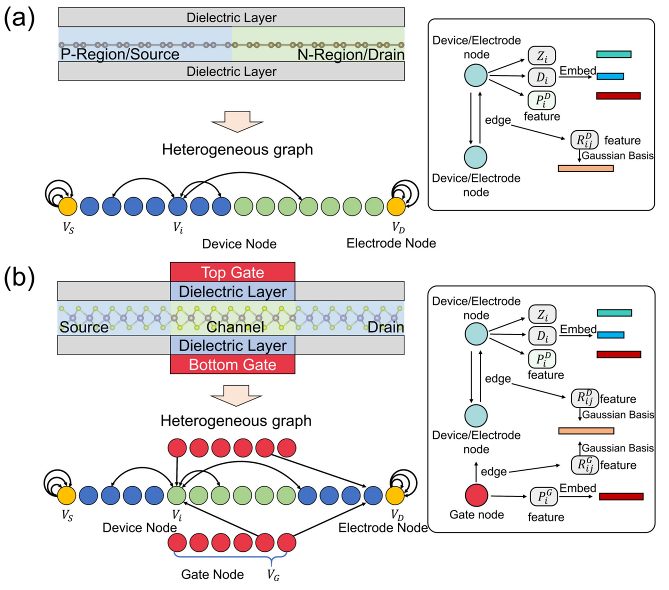

Fig. 2. Heterogeneous graph abstraction of devices. Yellow represents electrode nodes, blue and green represent device nodes, and red represents gate nodes. Electrode nodes and device nodes have features including: atomic type, doping, and potential, and these nodes are connected to each other by bidirectional edges. Gate nodes, which exclusively feature potential, connect to electrode/device nodes via unidirectional edges. Both bidirectional and unidirectional edges have a distance feature. Notably, electrode nodes have multiple self-pointing edges, characterizing the semi-infinite electrodes. (a) Device abstraction with two terminals. (b) Device abstraction with three terminals.

and potential (P^D). These device nodes are interconnected by bidirectional edges, indicating the possibility of atomic interaction. Each edge is characterized by a single feature: the distance between atoms (R^D).

The atoms at the ends of the device serve as the source and drain electrodes, which we treat in two cases.

1) *Neumann Boundary Condition:* When we apply the Neumann boundary condition at the source-drain electrodes in Poisson's equation, the potential at the boundary is continuous and its magnitude can vary. This is because the potential at the source-drain electrode atoms is influenced not only by the applied voltage (V_{DS}) but also by the internal atoms. In our approach, we treat electrode nodes and device nodes uniformly within the graph to preserve their interactions. However, we enhance the representation of electrode nodes by adding multiple self-loop edges. Additionally, we initialize the potential value for electrode nodes with the source-drain voltage (V_{DS}) in the input to the neural network. The significance of this initial potential feature is maintained through a residual connection. It is crucial to explain the rationale behind this approach. In NEGF calculations, the effect of semi-infinite contact electrodes is described using contact self-energy, often solved via recursive algorithms. The termination criterion for these recursive calculations is that the self-energy does not change by more than a small value (ϵ) when additional electrode atoms are considered. To represent this in our model, we emulate the physical scenario of the semi-infinite contact electrode by adding multiple self-loop edges to the electrode nodes. The number of these edges corresponds to the convergence criterion of the contact self-energy solution, treated as a hyperparameter in our approach.
2) *Dirichlet Boundary Condition:* When we apply the Dirichlet boundary condition at the source-drain electrodes in Poisson's equation, the potential at the boundary is fixed. In this treatment, we distinguish between electrode nodes and device nodes as two types of nodes within the graph, where the electrode nodes can influence the device nodes and vice versa. In addition to the source-drain electrodes, a three-terminal device includes gate electrodes, which influence the internal atoms through an electric field without direct electron exchange. The gate contact is typically incorporated into Poisson's equation using the Dirichlet boundary condition, without considering a gate leakage current. This approach allows the voltage applied to the gate to affect the internal atoms without influencing the gate itself. In our approach, we introduce a distinct type of node for the gate in the graph, separate from device nodes and source/drain electrode nodes. The number of gate nodes reflects the length of the gate. Since gate nodes represent an abstraction of gate contacts rather than real atoms, they are characterized solely by a constant potential (P^G), unlike the variable potential characteristic of a device node. The gate nodes are connected to the device and electrode nodes by unidirectional edges, determined by the distance between them (R^G).

We use the distance feature (R_D/R_G), which has been processed by the offset coefficient (Δ) and the scaling exponent (a)

$$R^{D/G} = \left(R_0^{D/G} + \Delta\right)^a. \tag{1}$$

As this relates to how the central node is influenced by other nodes at various distances in the global attention mechanism of the GFGNN, these parameters help us introduce a richer form of interaction, thereby increasing the network's expressiveness.

IV. Architecture of GFGNN

In device transport simulations within the NEGF formalism, the device Hamiltonian must be constructed using methods such as DFT, TB, k·p model, or EM. In this work, a two-band Dirac Hamiltonian is applied for device simulations based on graphene nanoribbon [35] and monolayer MoS_2 [36], [37]. The Hamiltonian is discretized onto real space grids [36], analogous to atoms in DFT or TB Hamiltonians. Consequently, the proposed GFGNN can be extended to calculations using DFT-NEGF and TB-NEGF. In k·p-based NEGF calculations, the channel Hamiltonian (H_0) remains constant, while the channel potential is updated self-consistently. This allows for a simplification in GFGNN-accelerated device simulations: predicting the self-consistent Hamiltonian of the open system reduces to predicting the potential distribution of the open system. The complete workflow is depicted in Fig. 3. To model

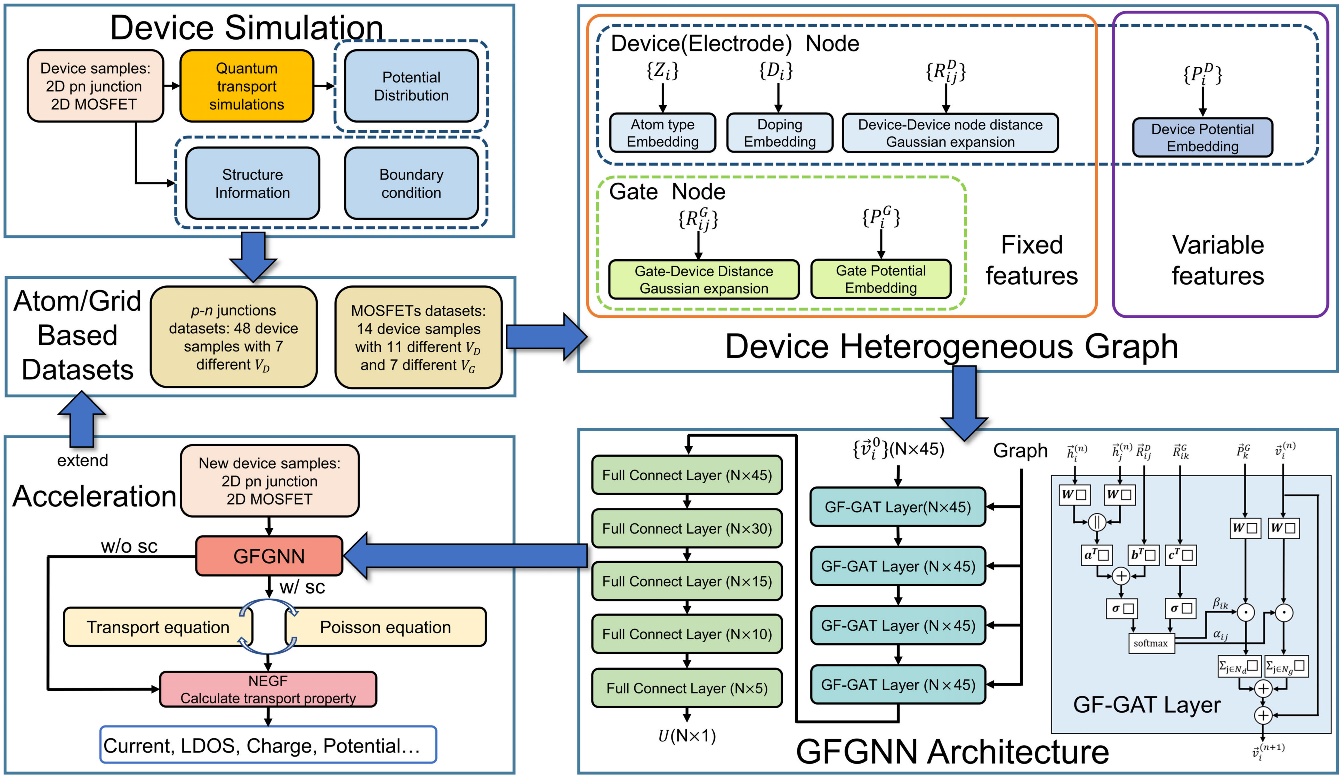

Fig. 3. Architecture and workflow of GFGNN: 1) device simulation—data obtained include potential distribution, structural information, and boundary conditions; 2) atom/grid-based datasets—consists of p-n junction datasets and MOSFET datasets; 3) device heterogeneous graph—defines device node features and gate node features, including both variable and fixed features; 4) GFGNN architecture—illustrates the information propagation mechanism of the GF-GAT layer with a schematic; and 5) acceleration of device simulation—demonstrates how the model enhances simulation efficiency.

devices like p-n junctions or MOSFETs, we preform quantum transport simulations and obtain the devices' structural information, boundary conditions, and potential distributions. The data are organized into atom- or grid-based datasets, then each device state is abstracted into a heterogeneous graph with nodes and edges. The features are transformed using embedding or Gaussian expansion, classifying nodes and edges by type. These graphs are trained using a GFGNN with global field graph attention network (GF-GAT) layers and full connected (FC) layers. Finally, we integrate the trained model into NEGF simulations, either accelerating self-consistent iterations without losing accuracy or replacing them entirely for faster computation, though with some accuracy trade-offs. This approach can also expand the dataset, enhancing the GFGNN's predictive power. Here, we will introduce two critical components of the workflow: 1) the device heterogeneous graph; and 2) the network core structure, the GF-GAT layer.

The device heterogeneous graph comprises numerous nodes and edges, each characterized by specific features. From a fundamental physics perspective, these features can be classified into fixed and variable types. For instance, atom type, doping, interatomic distance, and gate node potentials are fixed features, whereas the potentials of the device nodes are variable and indeterminate. These two types of features play distinct roles within the network: During network propagation, only the variable features undergo updates, and this process is driven by the interplay between both fixed and variable features.

The features must first be transformed into vectors that the network can learn. Depending on the characteristics of the features, we utilize either embedding or Gaussian basis expansion

$$\begin{aligned}\vec{Z}_i, \vec{D}_i, \vec{P}_i^D, \vec{P}_i^G &= \text{Embedding}\big(Z_i, D_i, P_i^D, P_i^G\big)\\ \vec{R}_{ij}^D, \vec{R}_{ij}^G &= \text{GaussExp}\big(R_{ij}^D, R_{ij}^G\big).\end{aligned} \tag{2}$$

Since the variable feature is updated during network propagation, we define a new variable feature $\vec{v}_i^{(n)}$ to distinguish it from the initial value, where i is the device node number and n is the number of network layers. We use the initial device node potential vector $\vec{P}_i^D$, which includes the source-drain voltage information, as the initial value of the variable feature (both $\vec{P}_i^D$ and $\vec{v}_i^{(n)}$ represent the device node potential features, but the former indicates the initial value and the latter represents the variable feature during network propagation)

$$\vec{v}_i^{(0)} = \vec{P}_i^D. \tag{3}$$

Then, the device graph and its features are fed into the GF-GAT layer, which we propose inspired by the Transformer's global attention mechanism [38] and exploits the simplicity of the GAT network for attention computation [39].

In the GF-GAT layer, the update of variable feature is based on an attention mechanism, and we utilized a message variable $\vec{m}_i^{(n+1)}$ to convey the network updates for $\vec{v}_i^{(n)}$, where N_d is the number of all device nodes and N_g is the number of all gate nodes, and we enable each device node to be aware of the effects of all device nodes (including itself) and gate nodes

$$\vec{m}_i^{(n+1)} = \sigma\left(\sum_{j\in\mathcal{N}_d}\alpha_{ij}\mathbf{W}_d\vec{v}_j^{(n)} + \sum_{j\in\mathcal{N}_g}\beta_{ij}\mathbf{W}_\mathbf{g}\vec{P}_j^G\right) \tag{4}$$

where α_{ij} is the attention coefficient between device nodes. Here, the fixed features of the nodes, and variable feature are the influencing factors of the interactions between the nodes and participate in the calculation of α_{ij} according to their respective characteristics

$$\begin{aligned}\alpha_{ij} &= \mathrm{softmax}\left(e_{ij}\right)\\ &= \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathcal{N}_d}\exp(e_{ik}) + \sum_{k\in\mathcal{N}_g}\exp(E_{ik})}\end{aligned} \tag{5}$$

$$e_{ij} = \mathrm{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{W}\vec{h}_i^{(n)}\|\mathbf{W}\vec{h}_j^{(n)}\right] + \mathbf{b}^T\left(\vec{R}_{ij}^D\right)\right) \tag{6}$$

$$\vec{h}_i^{(n)} = \left(\vec{Z}_i\,\|\,\vec{D}_i\,\|\,\vec{v}_i^{(n)}\right) \tag{7}$$

where β_{ij} is then the attention factor of the gate nodes to the device nodes. The distance feature between the two is used in the calculations, which, in combination with the gate potential multiplied during the feature update (4), reproduces the field effects in the gate contact

$$\begin{aligned}\beta_{ij} &= \mathrm{softmax}\left(E_{ij}\right)\\ &= \frac{\exp\left(E_{ij}\right)}{\sum_{k\in\mathcal{N}_d}\exp(e_{ik}) + \sum_{k\in\mathcal{N}_g}\exp(E_{ik})}\end{aligned} \tag{8}$$

$$E_{ij} = \mathrm{LeakyReLU}\left(\mathbf{c}^T\left(\vec{R}_{ij}^G\right)\right). \tag{9}$$

We aggregate the information $\vec{m}_i^{(n+1)}$ obtained from each GF-GAT layer with the initial input $\vec{v}_i^{(n)}$ of each layer to produce the output $\vec{v}_i^{(n+1)}$

$$\vec{v}_i^{(n+1)} = \vec{m}_i^{(n+1)} + \vec{v}_i^{(n)}. \tag{10}$$

After the propagation of information through the N GF-GAT layers, it will output the potential feature (variable feature) of each device node $v_i^{(N)}, i \in [1, \ldots, N_d]$. Each potential feature is a multidimensional vector that, after inputting several FC layers, outputs the final predicted device node potential u_i (scalar).

In the subsequent workflow, we add the potentials of the device nodes $U = \mathrm{diag}\{u_1, \ldots, u_{N_d}\}$ and the device's Hamiltonian H_0 to obtain the device's open-system self-consistent Hamiltonian, H

$$H = H_0 + U. \tag{11}$$

We then use H as the initial value for the self-consistent iterative Hamiltonian to accelerate convergence. With sufficiently high accuracy, we can also bypass the self-consistent step and directly input H into the subsequent nonself-consistent step for transport calculations, allowing us to obtain properties such as current, local density of states (LDOSs), charge, and potential. For detailed information on the neural network architecture and training specifics, please refer to the experimental setup in Table I. The network is implemented using PyTorch 2.4.1. It utilizes four GF-GAT layers for information propagation, each with input and output dimensions of 45. A high-dimensional mapping function is constructed with five FC layers, transitioning from input dimensions of [45, 30, 15, 10, 5] to output dimensions of [30, 15, 10, 5, 1]. The network is trained with a batch size of 5, 200 epochs, a learning rate of 0.005, using $L1$ Loss, and the Adam optimizer. The trained GFGNN, exemplified with MOSFETs, has 79 555 parameters and takes 639.19 s to train on an NVIDIA A100 GPU. The MOSFETs training set includes eight devices with 103 180 data points. For NEGF transport simulations of each MOSFET on a 72-core Intel Xeon CPU, the time taken ranges from 1804.25 to 8894.97 s when scanning V_G with a fixed V_D. It is evident that, even with the use of 72 CPU cores for parallel computation in numerical simulations, the network training time is significantly shorter than a single device simulation time. This demonstrates the model's considerable advantage in reducing overall simulation time.

TABLE I
EXPERIMENTAL SETUP

Model architecture and hyperparameters	Value
Number of GF-GAT layers	4
Dimensions of the input and output of the GF-GAT layer	[45,45,45,45,45]
Number of FC layers	5
Dimensions of the input and output of the FC layer	[45,30,15,10,5,1]
Batch size	5
Epoch number	200
Learning rate (Decays by a factor of 0.5 every 50 epochs)	0.005
Loss function	$L1$ Loss
Optimizer	Adam
Number of parameters of GFGNN (take MOSFET training set as an example)	79,555
Training time (trained on a Nvidia A100 GPU and take MOSFET training set as an example)	639.19 s
Training set (take MOSFET training set as an example)	8 device samples (Including 103,180 data points)
Simulation time (simulated on Intel Xeon CPU with 72-core for MOSFET datasets)	1804.25 - 8944.97 s

V. Device Structure and Datasets

To verify the effectiveness of our proposed method, we applied the entire workflow to representative two-terminal devices (p-n junctions) and three-terminal devices (MOSFETs). For these devices, quantum transport simulations are carried out using in-house developed simulator, by solving the Poisson and Schrödinger equations self-consistently. For each voltage point, the initial potential distribution is set to the final value from the previous iteration. During the self-consistent process, Anderson mixing is employed to update the potential distribution. The in-house developed simulator has been successfully applied to the research of Dirac-source field-effect transistors (DS-FETs) [36], cold-source FETs (CS-FETs) [40], [41], and other emerging devices [42], consistently producing reliable results.

The structure of the p-n junction is shown in the top left corner of Fig. 2(a). The dielectric layer has a thickness of 1 nm and a dielectric constant of 3.9 (SiO_2). The p- and n-region lengths are denoted as L_p and L_n. Both L_p and L_n are variables in our tests. The doping concentration in both the

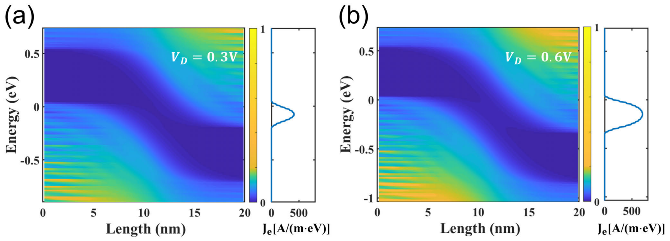

Fig. 4. LDOS and current density for the p-n junction with $L_n = 11.36$ nm and $L_p = 8.52$ nm. (a) $V_D = 0.3$ V and (b) $V_D = 0.6$ V. At this stage, the tunneling current in the p-n junction becomes dominant.

p- and n-regions is 2.86×10^{12} cm^{-2} (with different types). The channel material is GNR with a bandgap of 0.55 eV and the Hamiltonian is described using the Dirac model [35]. Notably, we use a material with a relatively small bandgap and keep the p-n junction under reverse bias in subsequent tests (not exhibiting rectifying characteristics). This approach evaluates GFGNN's applicability in quantum tunneling scenarios and validates its accuracy in capturing the intrinsic physics of quantum transport. As shown in Fig. 4, the tunnel current through the junction between the p-type valence band and n-type conduction band increases significantly as the bias voltage rises from $V_D = 0.3$ V to $V_D = 0.6$ V.

The MOSFET structure is depicted in the bottom left corner of Fig. 2(b). The dielectric layer has a thickness of 1 nm and a dielectric constant of 20 (HfO_2). The source, channel, and drain region lengths are denoted as L_S, L_C, and L_D, respectively. Here, the lengths of L_S and L_D are fixed at 7.1 nm, and L_C is a variable in our tests. The doping concentration in both the source and drain regions is 2.86×10^{13} cm^{-2}, and the channel region is undoped. The channel material is MoS_2 with a bandgap of 1.65 eV, and the Hamiltonian is described using the Dirac model [36], [37].

In our neural network prediction tests, we chose device length as the primary variable for structural variation because the most pressing issue in quantum transport simulations is the computational burden that increases with device size. If models trained on data from devices of limited sizes could predict the properties of both smaller and larger devices, it would significantly alleviate the issue. Of course, the model's ability to predict changes in other parameters is also crucial. Parameters such as atomic types, doping, and dielectric constants are explicitly represented in the GFGNN framework. With an appropriately constructed training set, the model can handle these variations as well.

VI. Application to P-N Junctions

The performance of a two-terminal p-n junction is only influenced by the source-drain voltage. We conducted the following two tests.

1) *Interpolation Capability Test:* To evaluate the interpolation capability of GFGNN in two-terminal devices, we used data from 30 p-n junctions with varying p-region lengths ($L_p = 4.97$–14.91 nm; step = 0.71 nm) and n-region lengths ($L_n = 4.97$–14.91 nm; step = 0.71 nm) as the training and validation sets. We then tested the model with data from three kinds of p-n junctions whose p-region and n-region lengths fell within this range but were not included in the training and validation sets. The p-n junction voltages varied from 0.0 to 0.6 V (step = 0.1 V). The training dataset was extensive, as each p-n junction included seven different voltage scenarios, with many atoms representing nodes in different environments within the device graph. In total, 48 880 data points were used for training.

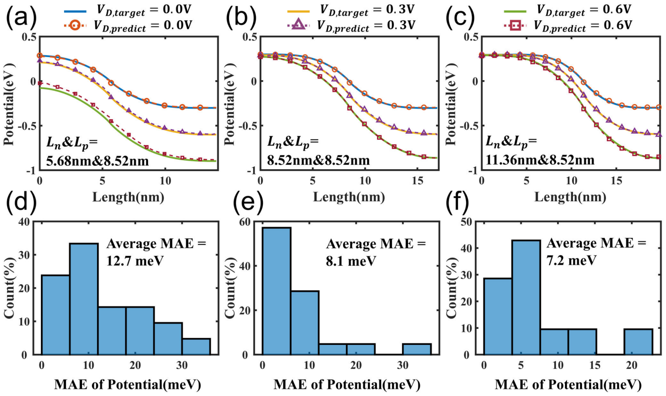

Fig. 5. Interpolation capability test of GFGNN in p-n junctions. Comparison between the predicted and target potentials of the p-n junctions for different n- and p-region lengths (L_n and L_p) of (a) 5.68 and 8.52 nm, (b) 8.52 and 8.52 nm, and (c) 11.36 and 8.52 nm, respectively, at $V_D =$ 0, 0.3, 0.6 V. (d)–(f) MAE distribution predicted for all voltage points for the cases in (a)–(c), where the average MAE is 12.7, 8.1, and 7.2 meV, respectively.

2) *Extrapolation Capability Test:* To evaluate the extrapolation capability of GFGNN in two-terminal devices, we fixed the p-region length at $L_p = 8.52$ nm and varied the n-region length from ($L_n = 5.68$–15.62 nm; step = 0.71 nm), resulting in a total of 15 p-n junctions data points for the training and validation sets. For the test set, we extended the n-region length to $L_n = 16.63$–17.75 nm. The p-n junction voltage varied from 0.0 to 0.6 V (step = 0.1 V). Our training dataset comprised 12 536 data points.

These tests allowed us to rigorously assess the ability of GFGNN to both interpolate and extrapolate in the context of two-terminal p-n junction devices.

Results of Test (1): In Fig. 5(a)–(c), we compare the potential predictions and numerical calculation results for p-n junctions with L_n and L_p lengths of 5.68 and 8.52 nm, 8.52 and 8.52 nm, and 11.36 and 8.52 nm, respectively. To illustrate the effect of varying source-drain voltage V_D on the potential distribution, we selected three cases with $V_D =$ 0, 0.3, and 0.6 V for comparison. As shown in the figures, the differences between the predicted and actual potentials are minimal. To quantify this difference, Fig. 5(d)–(f) presents the mean absolute error (MAE) distribution of the potentials at all voltage points. In all three devices, the MAE is predominantly below 10 meV, with average MAE values of 12.7, 8.1, and 7.2 meV, respectively.

Furthermore, in the case of nonheavily doped p-n junctions with short p-regions, the source side is susceptible to the applied voltage at the drain side, causing the potential to be

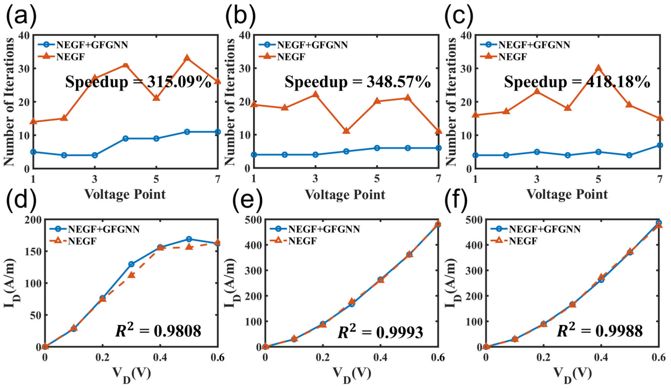

Fig. 6. Acceleration capability and computational accuracy of GFGNN-embedded NEGF in p-n junctions. (a)–(c) Comparison of the number of iterations per voltage point between conventional NEGF and GFGNN-embedded NEGF, showing acceleration rates of 315.09%, 348.57%, and 418.18%, respectively. (d)–(f) Comparison of I_D–V_D curves calculated by conventional NEGF and GFGNN-embedded NEGF for the above cases, with $R^2 = 0.9808$, 0.9993, and 0.9988, respectively.

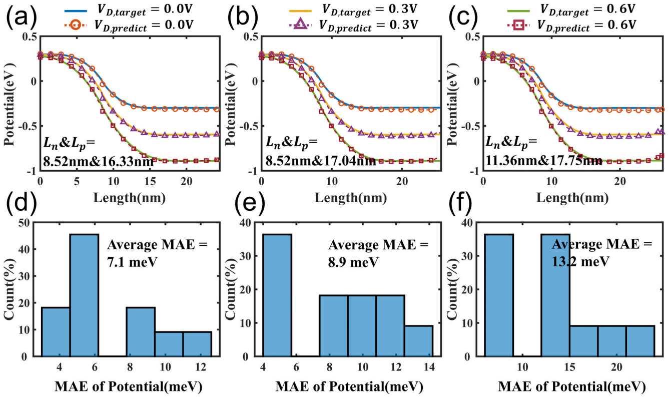

Fig. 7. Extrapolation capability test of GFGNN in p-n junctions. Comparison of the predicted and target potentials for p-n junctions with n- and p-region lengths (L_{n} and L_{p}) of (a) 8.52 and 16.33 nm, (b) 8.52 and 17.04 nm, and (c) 8.52 and 17.75 nm, respectively, at $V_D = 0$, 0.3, 0.6 V. (d)–(f) Distribution of MAE predicted for all voltage points in the cases shown in (a)–(c), with average MAE values of 7.1, 8.9, and 13.2 meV, respectively.

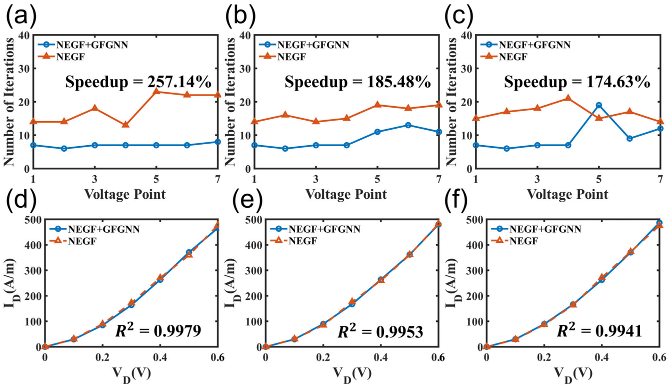

Fig. 8. Acceleration capability and computational accuracy of GFGNN-embedded NEGF in p-n junctions. (a)–(c) Comparison of the number of iterations per voltage point (V_D) between conventional NEGF and GFGNN-embedded NEGF for the cases of L_{n} and $L_{\text{p}} = 8.52$ and 16.33 nm, (b) 8.52 and 17.04 nm, and (c) 8.52 and 17.75 nm, with the acceleration = 257.14%, 185.48%, and 174.63%, respectively. (d)–(f) Comparison of I_D–V_D curves calculated by conventional NEGF and GFGNN-embedded NEGF in the above cases, with $R^2 = 0.9979$, 0.9953, and 0.9941, respectively.

nonfixed. This physical phenomenon is well captured by the GFGNN predictions, which show a high degree of agreement with the numerical calculations. This demonstrates the ability of GFGNN to accurately capture and learn the intrinsic physics of p-n junctions with varying lengths.

Next, we embedded the trained GFGNN into the NEGF computational framework to verify its acceleration effect. We employed GFGNN to predict a potential distribution as an initial value for each voltage point's self-consistent iteration. This approach aims to accelerate the process by reducing the number of iterations required for each voltage point. The comparison of the number of iterations required before and after acceleration at different voltage points is shown in Fig. 6(a)–(c). The results demonstrate that GFGNN significantly accelerates the iteration process at all voltage points, achieving speedups of 315.09%, 348.57%, and 418.18% in the three p-n junctions, respectively. Alongside this speed improvement, ensuring computational accuracy is crucial. We compared the I_D–V_D curves before and after acceleration in Fig. 6(d)–(f), which show high agreement with R^2 values of 0.9808, 0.9993, and 0.9988, respectively.

Results of Test (2): In Fig. 7(a)–(c), we compare the predicted and numerically calculated potentials for the p-n junctions with L_{n} and L_{p} values of 8.52 and 16.33 nm, 8.52 and 17.04 nm, and 8.52 and 17.75 nm, respectively. The three voltage scenarios, $V_D = 0$, 0.3, and 0.6 V, are also included. The figures illustrate that the discrepancy between the predicted and target values is minimal; however, this error increases gradually with the device length. This trend is more clearly depicted in Fig. 7(d)–(f), where the average MAE is 7.1, 8.9, and 13.2 meV as the device length increases, with errors predominantly concentrated around the source-drain electrodes. Our analysis suggests the following: 1) since the training data included only small-sized devices, the environment experienced by each atom increasingly deviates from the original training conditions as the device length increases. Despite this, GFGNN continues to demonstrate strong predictive capabilities, indicating its ability to extrapolate and learn intrinsic physics; 2) as the sample deviates further from initial observations, prediction errors in various physical formulations increase, a trend also observed in GFGNN; and 3) for large-sized device predictions, errors are mainly found near the source-drain electrodes, suggesting potential optimization opportunities in the parameter processing of electrode nodes. Embedding the trained GFGNN into the NEGF computational framework, as demonstrated in Fig. 8(a)–(c), resulted in acceleration factors of 257.14%, 185.48%, and 174.63%, respectively. In the I_D–V_D curves [Fig. 8(d)–(f)], there is a high agreement between NEGF and GFGNN + NEGF, with R^2 values of 0.9979, 0.9953, and 0.9941, respectively.

VII. Application to MOSFETs

In three-terminal devices, the performance is influenced not only by the source-drain voltage but also by the gate voltage. Due to the differences in the heterogeneous graph abstraction between two-terminal and three-terminal devices, as well as variations in network algorithm optimizations, it is essential to further validate the performance of GFGNN in three-terminal devices. This validation utilizes MoS_2-based MOSFETs. We conducted the following two tests.

1) *Extrapolation Capability Test:* To evaluate the extrapolation capability of GFGNN in a three-terminal device, we fixed the lengths of the source and drain regions at $L_S = L_D = 7.1$ nm, and varied the length of the channel region, L_C from 7.1 to 12.07 nm in steps of 0.71 nm. Data from a total of eight MOSFETs were used for training and validation. The channel region length L_C was then extended to a range of 12.78–14.20 nm for the test set. Unlike p-n junctions, MOSFETs require consideration of both the gate voltage V_G and the source-drain voltage V_D, adding additional dimensions to the data. Specifically, V_G varied from 0 to 0.6 V in steps of 0.1 V, and V_D varied from 0 to 0.5 V in steps of 0.05 V. The training data comprised a total of 103 180 data points.
2) *Skip Self-Consistent Step Test:* The GFGNN-predicted open-system Hamiltonian was used directly in the nonself-consistent step. In previous tests, we incorporate GFGNN-predicted results as initial values in the self-consistent iteration, achieving significant acceleration provided the predicted results are close to the final iteration values. In this approach, we input the predicted results from the test set directly into the nonself-consistent step to compute the transport properties. By bypassing the self-consistent iterative process entirely, this method maximizes acceleration. However, it is important to quantitatively assess the tradeoff in accuracy.
3) *Scalability Testing of Other Device Parameters—Considering Variations in the Gate Dielectric Constant:* In previous tests, our discussions on scalability focus primarily on device dimensions. To verify the compatibility of GFGNN with learning other device parameters, we use the gate dielectric constant as an example to demonstrate the general approach and process for making GFGNN adaptable to a wider range of device parameters.

A. Result of Test (1)

In Fig. 9(a)–(c), we compare the predicted and target values of the potential distribution at $V_D = 0.25$ V for channel lengths $L_C = 12.78$, 13.49, and 14.20 nm. The curves show high agreement, with minimal error differences across other values of V_D. In Fig. 9(d)–(f), we present the distribution of the MAE between predicted and target values for the three devices, with average MAE values of 2.1, 2.6, and 3.3 meV, respectively. These values are in the meV range and are even lower than those observed in two-terminal devices. Notably, the potential distribution predicted by GFGNN shows a bump in potential between the source and the channel, which is absent between the drain and the channel. This bump reflects the drain-induced barrier lowering (DIBL) effect, highlighting GFGNN's capability to capture and learn from intrinsic device physics. Next, we embed the trained GFGNN into the NEGF computational framework. The good acceleration effect of the GFGNN at each voltage point is demonstrated in Fig. 10(a)–(c), and the acceleration is 184.85%, 189.74%, and 182.22%, respectively. In the I_D–V_G curves [Fig. 10(d)–(f)], NEGF and GFGNN + NEGF are also in high agreement with $R^2 = 0.9992$, 1.0000, and 0.9999, respectively.

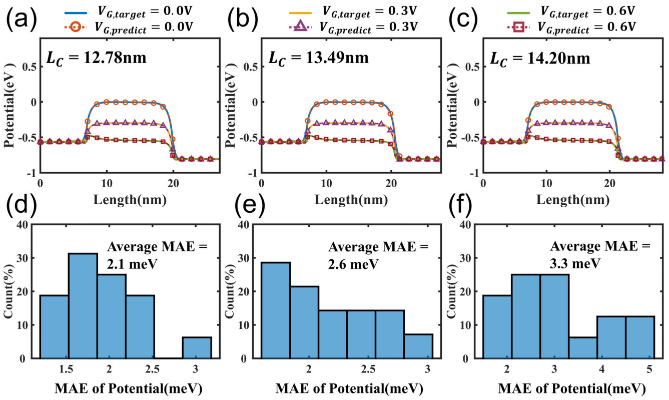

Fig. 9. Extrapolation capability test of GFGNN in MOSFETs. Comparison of the predicted and target potentials for MOSFETs with fixed source/drain lengths (L_S/L_D) of 7.1 nm and channel lengths (L_C) of (a) 12.78 nm, (b) 13.94 nm, and (c) 14.20 nm. The cases include $V_D = 0.25$ V and $V_G = 0$, 0.3, and 0.6 V. (d)–(f) Distribution of MAEs for all voltage points in the cases shown in (a)–(c), with average MAE values of 2.1, 2.6, and 3.3 meV, respectively.

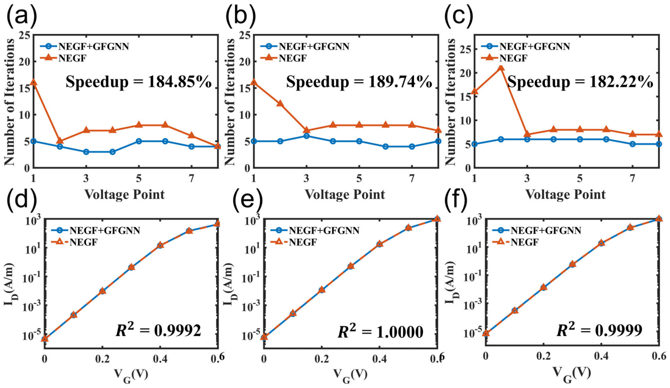

Fig. 10. Acceleration capability and computational accuracy of GFGNN-embedded NEGF in MOSFETs. (a)–(c) Comparison of the number of iterations per voltage point (V_G) between conventional NEGF and GFGNN-embedded NEGF for the cases of $L_C = 12.78$, 13.94, and 14.20 nm, with the acceleration = 184.85%, 189.74%, and 182.22%, respectively. (d)–(f) Comparison of I_D–V_G curves calculated by conventional NEGF and GFGNN-embedded NEGF in the above cases, with $R^2 = 0.9992$, 1.0000, and 0.9999, respectively.

To better demonstrate the robustness of GFGNN's predictions at smaller scales and its scalability at larger scales, we conducted further tests on the trained model. First, we predicted the potential distribution for a MOSFET with $L_C = 5.68$ nm at $V_D = 0.25$ V, as shown in Fig. 11(a). Even in devices with significant quantum effects, GFGNN exhibited high predictive accuracy, with an average MAE of only 3.8 meV. Subsequently, we predicted the potential distributions for MOSFETs with $L_C = 17.75$ and 21.30 nm at $V_D = 0.25$ V, as shown in Fig. 11(b) and (c). In these larger scale devices, the average MAEs are 4.9 and 8.2 meV, respectively.

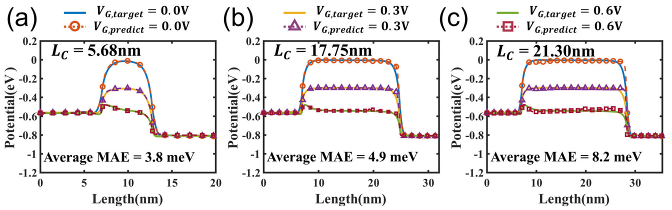

Fig. 11. Testing the robustness of GFGNN's predictions at smaller scales and its scalability at larger scales. Comparison of the predicted and target potentials for MOSFETs with fixed source/drain lengths (L_S/L_D) of 7.1 nm and channel lengths (L_C) of (a) 5.68 nm, (b) 17.75 nm, and (c) 21.30 nm. The cases include $V_D = 0.25$ V and $V_G = 0$, 0.3, and 0.6 V. The average MAE of the predictions are (a) 3.8 meV, (b) 4.9 meV, and (c) 8.2 meV, respectively.

It is noteworthy that while the range of L_C in the training set is from 7.1 to 12.07 nm, GFGNN is able to accurately predict within the range from 12.07 to 21.30 nm, highlighting its strong scalability due to the comprehensive learning of the device's intrinsic physics. This also demonstrates that the model, with its enhanced learning capabilities, can mitigate the need for extensive training data.

B. Result of Test (2)

Unlike previous tests, we directly use the Hamiltonian predicted by GFGNN as input for the nonself-consistent step. Given the high accuracy of GFGNN predictions, the computed transport properties are comparable in accuracy to those obtained from the conventional computational framework, while achieving significantly improved computational speed. Fig. 12(a)–(c) shows a comparison between the NEGF and GFGNN + NEGF I_D–V_G curves without self-consistency, with $R^2 = 0.9997$, 0.9961, and 0.9996, respectively. Fig. 12(d)–(f) displays the LDOS and error distribution for both methods. These results highlight the exceptionally high accuracy of the GFGNN + NEGF method for transport calculations. To quantify the speedup, we compare the computation time based on the number of times the transport equation is solved. In the calculations shown in Fig. 12(a)–(c), the conventional NEGF method requires 68, 81, and 89 steps, respectively, whereas the nonself-consistent NEGF + GFGNN approach requires only seven steps (corresponding to seven voltage points), resulting in speedups of 971.43%, 1157.14%, and 1271.43%.

C. Result of Test (3)

For more complex systems with additional parameters, the key is to generate datasets for these extra device parameters and construct heterogeneous graphs accordingly. Specifically, in this case, we add the gate dielectric constant as a feature (ϵ^G) of the edge between the device node and the electrode node to the network information propagation. To achieve this, we incorporate a Gaussian basis expansion with respect to the gate dielectric constant and modify (9). The results are presented in (12) and (13). And then we fixed the device size constant ($L_C = 11.36$ nm), varied the dielectric constant of the gate by 5, 10, 15, 20, and scanned the source/drain voltage $V_{\text{DS}} = 0$ to 0.5 V (step = 0.05 V) and gate voltage $V_{\text{DS}} = 0$ to 0.6 V (step = 0.1 V). The dataset is divided into training set, validation set, and test set (0.8/0.1/0.1). After training, the average MAE on the test set is 4.2 meV. Further, we predict the potential distribution of MOSFETs with the Al_2O_3 gate dielectric layer (dielectric constant = 9), and the average MAE is 11.9 meV. The results are shown in Fig. 13

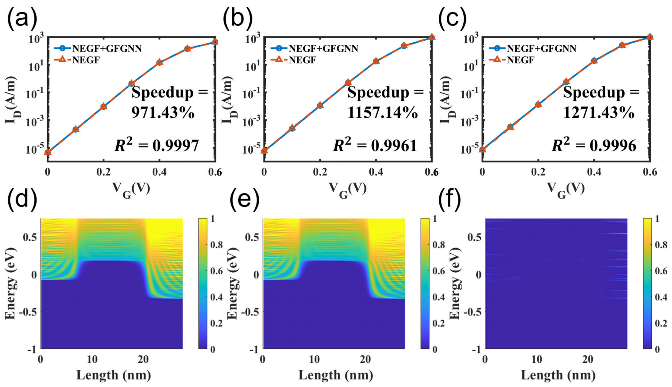

Fig. 12. Acceleration capability and computational accuracy of GFGNN-embedded NEGF (skipping self-consistent iteration) in MOSFETs. (a)–(c) Comparison of I_D–V_G curves between conventional NEGF and GFGNN-embedded NEGF for channel lengths L_C of 12.78, 13.94, and 14.20 nm, with $R^2 = 0.9997$, 0.9961, and 0.9996, respectively. The acceleration achieved is 971.43%, 1157.14%, and 1271.43%, respectively. (d) LDOS calculated by conventional NEGF. (e) LDOS calculated by GFGNN-embedded NEGF. (f) Error distribution.

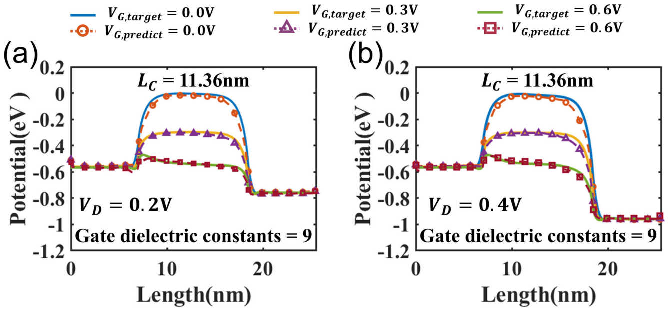

Fig. 13. Scalability testing of the gate dielectric constant: a comparison between the predicted potential distribution and the target values when the gate dielectric layer is Al_2O_3 (dielectric constant = 9). (a) $V_D = 0.2$ V and (b) $V_D = 0.4$ V.

$$\vec{\epsilon}_{ij}^{G} = \text{GaussExp}\left(\epsilon_{ij}^{G}\right) \tag{12}$$

$$E_{ij} = \text{LeakyReLU}\left(\mathbf{c}^{T}\left(\vec{R}_{ij}^{G}\right) + \mathbf{d}^{T}\left(\vec{\epsilon}_{ij}^{G}\right)\right). \tag{13}$$

VIII. Conclusion

Our study utilizes ML for quantum transport calculations and examines the differences and connections between confined/periodic systems and open systems. Specifically, we propose an attention-based GFGNN that describes field effects and dynamic features and directly predicts the self-consistent potential distribution of a device based on its structure and electrode voltages. We introduce heterogeneous graph representations for both two- and three-terminal devices, accommodating various boundary conditions (Neumann and Dirichlet). These representations preserve device features and are well-suited for GFGNN inputs.

We evaluate GFGNN's interpolation and extrapolation capabilities using GNR-based p-n junctions (two-terminal) and MoS_2-based MOSFETs (three-terminal), demonstrating high prediction accuracy. After integrating into the NEGF computational framework, we achieve computational acceleration between 174.63% and 418.18% for self-consistent calculations. In the MOSFET tests, the high prediction accuracy of GFGNN allowed us to bypass the self-consistent step, using the predicted potential distribution directly in the nonself-consistent step to compute transport properties. This approach achieved an acceleration of 971.43%–1271.43% while maintaining accuracy comparable to traditional NEGF computations.

Our strategy preserves the Hamiltonian's constant nature in confined/periodic systems while focusing on electrode-induced effects. This facilitates NEGF-related self-consistent loops and links naturally with existing efforts to predict confined/periodic-system Hamiltonians [25], [26]. By leveraging predictive models for estimating confined/periodic-system Hamiltonians, we can use our GFGNN to obtain the corresponding open-system Hamiltonian under applied voltages, ultimately computing transport properties with ab-initio accuracy.

Acknowledgment

The authors would like to thank the high-performance computing platform of Peking University for computation facilities.

References

[1] A. Razavieh, P. Zeitzoff, and E. J. Nowak, "Challenges and limitations of CMOS scaling for FinFET and beyond architectures," *IEEE Trans. Nanotechnol.*, vol. 18, pp. 999–1004, Sep. 2019, doi: 10.1109/TNANO.2019.2942456.

[2] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature Electron.*, vol. 1, no. 8, pp. 442–450, Aug. 2018, doi: 10.1038/s41928-018-0117-x.

[3] C. Qiu et al., "Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches," *Science*, vol. 361, no. 6400, pp. 387–392, Jul. 2018, doi: 10.1126/science.aap9195.

[4] H. Amrouch, G. Pahwa, A. D. Gaidhane, J. Henkel, and Y. S. Chauhan, "Negative capacitance transistor to address the fundamental limitations in technology scaling: Processor performance," *IEEE Access*, vol. 6, pp. 52754–52765, 2018, doi: 10.1109/ACCESS.2018.2870916.

[5] Y. Liu, X. Duan, H.-J. Shin, S. Park, Y. Huang, and X. Duan, "Promises and prospects of two-dimensional transistors," *Nature*, vol. 591, no. 7848, pp. 43–53, Mar. 2021, doi: 10.1038/s41586-021-03339-z.

[6] S. Das et al., "Transistors based on two-dimensional materials for future integrated circuits," *Nature Electron.*, vol. 4, no. 11, pp. 786–799, Nov. 2021, doi: 10.1038/s41928-021-00670-1.

[7] S. Datta, "The non-equilibrium Green's function (NEGF) formalism: An elementary introduction," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2002, pp. 703–706, doi: 10.1109/IEDM.2002.1175935.

[8] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[9] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly," *High-Confidence Comput.*, vol. 4, no. 2, Jun. 2024, Art. no. 100211, doi: 10.1016/j.hcc.2024.100211.

[10] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

[11] C.-T. Tung, M.-Y. Kao, and C. Hu, "Neural network-based and modeling with high accuracy and potential model speed," *IEEE Trans. Electron Devices*, vol. 69, no. 11, pp. 6476–6479, Nov. 2022, doi: 10.1109/TED.2022.3208514.

[12] G. Qi et al., "The device and circuit level benchmark of Si-based cold source FETs for future logic technology," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 3483–3489, Jun. 2022, doi: 10.1109/TED.2022.3164372.

[13] H. Xu, W. Gan, S. Guo, S. Zhang, and Z. Wu, "Machine learning-based compact modeling of silicon cold source field-effect transistors," *IEEE Trans. Nanotechnol.*, vol. 23, pp. 615–621, 2024, doi: 10.1109/TNANO.2024.3442476.

[14] L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, "Data-driven materials science: Status, challenges, and perspectives," *Adv. Sci.*, vol. 6, no. 21, Nov. 2019, Art. no. 1900808, doi: 10.1002/advs.201900808.

[15] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," *npj Comput. Mater.*, vol. 3, no. 1, pp. 1–13, Dec. 2017, doi: 10.1038/s41524-017-0056-5.

[16] T. Wu and J. Guo, "Speed up quantum transport device simulation on ferroelectric tunnel junction with machine learning methods," *IEEE Trans. Electron Devices*, vol. 67, no. 11, pp. 5229–5235, Nov. 2020, doi: 10.1109/TED.2020.3025982.

[17] S. Souma and M. Ogawa, "Neural network model for implementation of electron–phonon scattering in nanoscale device simulations based on NEGF method," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Dallas, TX, USA, Sep. 2021, pp. 56–59, doi: 10.1109/SISPAD54002.2021.9592557.

[18] P. Aleksandrov, A. Rezaei, N. Xeni, T. Dutta, A. Asenov, and V. Georgiev, "Fully convolutional generative machine learning method for accelerating non-equilibrium Green's function simulations," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Kobe, Japan, Sep. 2023, pp. 169–172, doi: 10.23919/sispad57422.2023.10319587.

[19] S.-C. Han, J. Choi, and S.-M. Hong, "Acceleration of semiconductor device simulation with approximate solutions predicted by trained neural networks," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5483–5489, Nov. 2021, doi: 10.1109/TED.2021.3075192.

[20] Y. Lim and M. Shin, "A novel machine-learning based mode space method for efficient device simulations," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Kobe, Japan, Sep. 2023, pp. 165–168, doi: 10.23919/sispad57422.2023.10319586.

[21] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Phys. Rev. Lett.*, vol. 120, no. 14, Apr. 2018, Art. no. 145301, doi: 10.1103/physrevlett.120.145301.

[22] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," *Nature*, vol. 624, no. 7990, pp. 80–85, Dec. 2023, doi: 10.1038/s41586-023-06735-9.

[23] Z. Deng, C. Chen, X.-G. Li, and S. P. Ong, "An electrostatic spectral neighbor analysis potential for lithium nitride," *npj Comput. Mater.*, vol. 5, no. 1, pp. 1–8, Jul. 2019, doi: 10.1038/s41524-019-0212-1.

[24] X.-G. Li, C. Hu, C. Chen, Z. Deng, J. Luo, and S. P. Ong, "Quantum-accurate spectral neighbor analysis potential models for Ni-Mo binary alloys and fcc metals," *Phys. Rev. B, Condens. Matter*, vol. 98, no. 9, Sep. 2018, Art. no. 094104, doi: 10.1103/physrevb.98.094104.

[25] H. Li et al., "Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation," *Nature Comput. Sci.*, vol. 2, no. 6, pp. 367–377, Jun. 2022, doi: 10.1038/s43588-022-00265-6.

[26] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, "General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian," *Nature Commun.*, vol. 14, no. 1, May 2023, Art. no. 2848, doi: 10.1038/s41467-023-38468-8.

[27] M. F. Kasim and S. M. Vinko, "Learning the exchange-correlation functional from nature with fully differentiable density functional theory," *Phys. Rev. Lett.*, vol. 127, no. 12, Sep. 2021, Art. no. 126403, doi: 10.1103/physrevlett.127.126403.

[28] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport," *Phys. Rev. B, Condens. Matter*, vol. 65, no. 16, Mar. 2002, Art. no. 165401, doi: 10.1103/physrevb.65.165401.

[29] M. P. Anantram and A. Svizhenko, "Multidimensional modeling of nanotransistors," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2100–2115, Sep. 2007, doi: 10.1109/TED.2007.902857.

[30] H. Carrillo-Nuñez, C. Medina-Bailón, V. P. Georgiev, and A. Asenov, "Full-band quantum transport simulation in the presence of hole-phonon interactions using a mode-space $k \cdot p$ approach," *Nanotechnology*, vol. 32, no. 2, Jan. 2021, Art. no. 020001, doi: 10.1088/1361-6528/abacf3.

[31] J. Guo, S. Datta, M. Lundstrom, and M. P. Anantam, “Toward multiscale modeling of carbon nanotube transistors,” *Int. J. Multiscale Comput. Eng.*, vol. 2, no. 2, pp. 257–276, 2004, doi: 10.1615/intjmultcompeng.v2.i2.60.

[32] F. A. Faber et al., “Prediction errors of molecular machine learning models lower than hybrid DFT error,” *J. Chem. Theory Comput.*, vol. 13, no. 11, pp. 5255–5264, Nov. 2017, doi: 10.1021/acs.jctc.7b00577.

[33] X. Wang et al., “Heterogeneous graph attention network,” in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032, doi: 10.1145/3308558.3313562.

[34] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, “Heterogeneous graph neural network,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 793–803, doi: 10.1145/3292500.3330961.

[35] S.-K. Chin, D. Seah, K.-T. Lam, G. S. Samudra, and G. Liang, “Device physics and characteristics of graphene nanoribbon tunneling FETs,” *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 3144–3152, Nov. 2010, doi: 10.1109/TED.2010.2065809.

[36] F. Liu, C. Qiu, Z. Zhang, L.-M. Peng, J. Wang, and H. Guo, “Dirac electrons at the source: Breaking the 60-mV/decade switching limit,” *IEEE Trans. Electron Devices*, vol. 65, no. 7, pp. 2736–2743, Jul. 2018, doi: 10.1109/TED.2018.2836387.

[37] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, “Coupled spin and valley physics in monolayers of MoS_2 and other group-VI dichalcogenides,” *Phys. Rev. Lett.*, vol. 108, no. 19, May 2012, Art. no. 196802, doi: 10.1103/physrevlett.108.196802.

[38] A. Vaswani et al., “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.

[39] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lió, and Y. Bengio, “Graph attention networks,” *Stat*, vol. 1050, no. 20, pp. 1–12, 2017.

[40] F. Liu et al., “First principles simulation of energy efficient switching by source density of states engineering,” in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2018, pp. 33.2.1–33.2.4, doi: 10.1109/IEDM.2018.8614597.

[41] H. Zhou et al., “Quantum transport simulations of sub-60-mV/decade switching of silicon cold source transistors,” *IEEE Trans. Electron Devices*, vol. 71, no. 4, pp. 2781–2788, Apr. 2024, doi: 10.1109/TED.2024.3370532.

[42] X. Xie, Z. Wang, X. Liu, and F. Liu, “Ternary cold source transistors for multivalue logic applications,” *Phys. Rev. Appl.*, vol. 22, no. 1, Jul. 2024, Art. no. 014053, doi: 10.1103/physrevapplied.22.014053.