electronics

Article

ResNet Modeling for 12 nm FinFET Devices to Enhance DTCO Efficiency

Yiming Huang [1,2], Bin Li [1,*], Zhaohui Wu [1] and Wenchao Liu [2]

[1] School of Microelectronics, South China University of Technology, Guangzhou 510641, China
[2] Primarius Technologies Co., Ltd., Building C, No. 888, Huanhu Xi Er Road, Lingang New Area, Pilot Free Trade Zone of China, Shanghai 430048, China
* Correspondence: phlibin@scut.edu.cn

Abstract: In this paper, a deep learning-based device modeling framework for design-technology co-optimization (DTCO) is proposed. A ResNet surrogate model is utilized as an alternative to traditional compact models, demonstrating high accuracy in both single-task (I–V or C–V) and multi-task (I–V and C–V) device modeling. Moreover, transfer learning is applied to the ResNet model, using the BSIM-CMG compact model for a 12 nm FinFET SPICE model as the pre-trained source. Through this approach, superior modeling accuracy and faster training speed are achieved compared to a ResNet surrogate model initialized with random weights, thereby meeting the rapid and efficient demands of the DTCO process. The effectiveness of the ResNet surrogate model in circuit simulation for 12 nm FinFET devices is demonstrated.

Keywords: design-technology co-optimization; FinFET; BSIM-CMG; SPICE model; ResNet surrogate model; transfer learning

Citation: Huang, Y.; Li, B.; Wu, Z.; Liu, W. ResNet Modeling for 12 nm FinFET Devices to Enhance DTCO Efficiency. *Electronics* **2024**, *13*, 4040. https://doi.org/10.3390/electronics13204040

Academic Editor: Elias Stathatos

Received: 12 September 2024
Revised: 10 October 2024
Accepted: 11 October 2024
Published: 14 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As device dimensions continue to shrink, traditional device manufacturing and circuit design processes face escalating challenges. The emergence of new structures [1,2], materials [3,4], mechanisms [5,6], and approaches [7] for devices has introduced significant difficulties for the rapid growth of the semiconductor industry based on Moore's law. A novel methodology is urgently needed to reduce the cost and time-to-market of next-generation devices. Design-technology co-optimization (DTCO) aims to continuously improve performance and power consumption by closely integrating device, process, and circuit design, optimizing each step even as technological advancements decelerate [8]. In the post-Moore's Law era, DTCO has become increasingly prominent [9].

Simulation Program with Integrated Circuit Emphasis (SPICE) models are an important link in the device-to-circuit chain. Rapid and extensive iterations are vital to DTCO methodology to keep up with the fast-evolving technology landscape. However, transitioning from compact models to SPICE models is time-consuming, often taking several weeks [10,11]. This time lag poses a significant bottleneck in the design cycle, necessitating more efficient modeling techniques.

Fin Field-Effect Transistor (FinFET) devices have become mainstream in advanced process nodes due to their superior performance [12,13]. However, the three-dimensional structure and complexity of FinFETs pose significant challenges for compact modeling. Traditional physics-based and empirical compact models, such as Berkeley Short-channel Insulated Gate Field Effect Transistor Model-Common Multi-Gate (BSIM-CMG), require extensive manual parameter fitting, which is labor-intensive and time-consuming [14,15]. Recently, machine learning methods have opened new avenues for FinFET modeling. In 2021, Wang J et al. [16] proposed an Artificial Neural Network (ANN)-based FinFET modeling method using limited Technology Computer Aided Design (TCAD) data. In

the same year, K. Xia [17] presented a method to create corner models for lookup table-based Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) models through input–output mapping. In 2022, M. Kao et al. [18] introduced an ANN-assisted physics-driven approach for MOSFET modeling. In the same year, Chien-Ting Tung et al. [19] demonstrated a neural network-based device modeling framework for advanced FETs, showing potential in accelerating circuit simulation speed compared to BSIM. In 2023, Hyunjoon Jeong et al. [20] developed an expandable ANN-based compact model for emerging transistors. In the same year, Jiahao Wei et al. [21] introduced an ANN-based MOSFET model for 0.18 μm analog processes, using novel data preprocessing and Latin hypercube sampling to reduce training data while outperforming the BSIM4 model in fitting and simulations. In 2024, Ya-Shu Yang et al. [22] presented a physical-based ANN framework for Gate-All-Around (GAA) MOSFETs. In the same year, Ziyao Yang et al. [23] proposed a graph-based compact model (GCM) using graph neural networks for efficient parameter extraction in 12 nm FinFETs.

However, existing machine learning modeling methods have limitations. The fitting and generalization capabilities of ANN models are limited, making it difficult to accurately describe the strong nonlinear characteristics of FinFET devices. Models trained from scratch often require large amounts of data and time, challenging the fast iteration demands of the industry. Existing methods mainly focus on fitting single characteristics, lacking unified modeling of multiple characteristics.

To overcome these limitations, a compact modeling method for FinFETs based on Residual Network (ResNet) is proposed. By introducing residual connections, ResNet can effectively extract and learn multi-scale features from data, demonstrating strong fitting capabilities for highly nonlinear functions. We introduce an innovative approach using ResNet into FinFET compact modeling and conduct pre-training and transfer learning based on the BSIM-CMG model, improving modeling efficiency and accuracy. Moreover, the ResNet surrogate model to circuit simulation is applied, validating the effectiveness of the proposed modeling method in the DTCO process.

The rest of this paper is organized as follows. Section 2 provides a comprehensive overview of the data preparation and preprocessing steps. Section 3 presents the training and testing procedures of the ResNet model. Section 4 discusses the results and analysis of the ResNet model's performance in both single-task and multi-task device modeling. Moreover, a 13-stage Ring Oscillator (RO) circuit based on FinFET devices was built to verify the accuracy of the ResNet model as compared to the commercial SPICE model with High-Speed SPICE (Hspice) simulation results. Section 5 discussions on the ResNet model's improvements and future work directions. Section 6 concludes the paper by summarizing the key findings and potential applications.

2. Data Preparation and Pre-Processing

In this study, extensive simulations of FinFET devices were conducted, focusing on critical device parameters and the corresponding simulation curves. A high-dimensional dataset was then constructed, which included channel length (L), channel width (W), number of fins (NFIN), temperature (Temp), gate-source voltage (Vgs), drain-source voltage (Vds), drain-source current (Ids), gate-source capacitance (Cgs), gate-drain capacitance (Cgd), and gate-bulk capacitance (Cgb), as listed in Table 1.

Table 1. Dataset parameter ranges and step sizes.

Parameter	Range	Step Sizes
L (nm)	16 to 240	4
W (nm)	10 to 1000	10
NFIN	1 to 20	1
Temp (°C)	−25 to 125	50
Vgs (V)	0 to 2	0.2
Vds (V)	0 to 2	0.05

Channel Length (L) and Channel Width (W) were identified as fundamental geometric dimensions that directly influence the device's scaling properties, threshold voltage, and drive current. Channel length scaling was recognized as critical for improving transistor speed and reducing power consumption, while channel width was observed to affect the current-carrying capability of the device.

The effective width of the FinFET was determined by the number of fins (NFIN), which directly impacts the drive strength and overall performance of the device. More fins typically result in higher drive current, making this a key design parameter.

Temperature (Temp) variations were found to significantly affect semiconductor performance, influencing carrier mobility and leakage currents. By including temperature as a parameter, the modeling of FinFET behavior under different operating conditions was enabled.

Gate-Source Voltage (Vgs) and drain-source voltage (Vds) were considered essential bias conditions that control the switching behavior of the FinFET. Vgs was observed to affect the inversion layer formation and control the on/off state of the device, while Vds was found to influence the current flow through the channel.

Drain-source current (Ids) was identified as the primary output characteristic and was determined to be a direct function of the applied voltages (Vgs and Vds) and the device's physical dimensions. It reflects the device's ability to conduct electricity in the on-state.

Gate-source capacitance (Cgs), gate-drain capacitance (Cgd), and gate-bulk capacitance (Cgb) were considered critical for understanding the switching speed and dynamic behavior of the device. These capacitances were found to affect the charging and discharging times during switching events, which are crucial for high-speed applications.

Moreover, the choice of these output characteristics, particularly Ids, Cgs, Cgd, and Cgb, was motivated by their suitability for packaging into scalable lookup table models. These models were intended to be implemented in Verilog-A syntax, enabling circuit-level simulations that accurately reflect the behavior of FinFET devices under various operating conditions. The use of lookup tables was recognized for allowing efficient interpolation and fast simulation, which is essential for validating device performance in large-scale circuit designs.

Therefore, the combination of these parameters was selected to provide a comprehensive representation of FinFET's electrical and physical characteristics. This approach was employed to ensure that a robust surrogate model could be developed, capable of accurately predicting device behavior, while also facilitating subsequent circuit simulation and verification using Verilog-A.

The Nano High-Speed SPICE (NanoSpice) simulator was used to simulate two different models. The BSIM-CMG compact model produced 29,000 files, each containing 451 data points with a total memory usage of 1.32 GB. The 12 nm FinFET SPICE model generated 24,000 files, each containing 451 data points with a total memory usage of 1.09 GB. These simulation datasets were used in our proposed deep learning-based device modeling framework for design-technology co-optimization (DTCO).

The deep learning-based device modeling framework was built on the Linux operating system. The server was configured with an AMD EPYC 7H12 64-Core Processor (AMD, Santa Clara, CA, USA) running at 1.5–2.6 GHz, an NVIDIA GeForce RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA), 2044 GiB of memory, and a 21 TB SSD (Samsung Electronics, Suwon, Republic of Korea). The deep learning algorithm was implemented using a Jupyter Notebook within the Anaconda environment. The Python version used was 3.10.12, and the PyTorch version was 2.0.1+cu117.

Data augmentation was performed based on the physical information of the devices, incorporating the physical and electrical characteristics into the input data via data augmentation, effectively embedding physical information into the model.

Min–max scaling was applied to the input data (L, W, NFIN, Temp, Vgs, Vds, and data augmentation) to reduce the magnitude differences between features in the high-dimensional dataset. This scaling accelerates model convergence and prevents certain

features from disproportionately influencing the loss function. For the target variables (Ids, Cgs, Cgd, and Cgb), logarithmic transformation was used to compress magnitude differences, and small constants were added to avoid infinite negative values from zero inputs. The dataset preprocessing is presented in Figure 1. The dataset was randomly split into training, validation, and test sets with a ratio of 7:2:1.

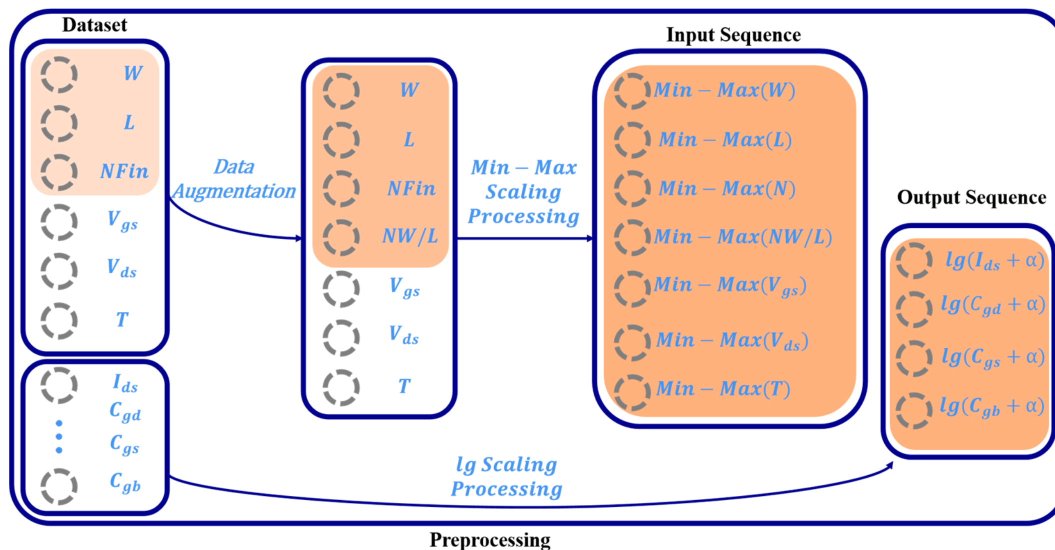


Figure 1. Preprocessing of the dataset.

3. Train and Test

It is known that a fully connected ANN structure is used in most deep learning-based surrogate models for semiconductor device use. While an ANN with sufficiently broad neuron architecture can accurately represent models, shallow networks with limited feature parameters struggle to maintain accuracy and generalization. These limitations lead to overfitting of the training data and inadequate approximation capabilities, necessitating deeper networks to manage the complexities of new semiconductor device structures. Although deep networks like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers have achieved significant success in image and natural language processing, modeling semiconductor devices aims to approximate real computational scenarios. Therefore, a fully connected ResNet structure was adopted to ensure modeling accuracy.

ResNet in this work incorporates traditional residual connections and varying neuron counts across hidden layers. Unlike conventional fully connected ResNets, where residual connections are added at every layer, residuals were strategically introduced at specific layers in our design. This selective approach reduces the number of residual summations, allowing for more efficient gradient computation and helping to control the complexity of the model. The varying number of neurons in different layers enables the model to scale and adapt at various levels, enhancing its ability to capture multi-level features more effectively.

More specifically, residual connections were applied after every two hidden layers, as shown in Figure 2.

Residual connections were added at strategic points, specifically after every two hidden layers, to maintain gradient flow through the network and reduce the risk of vanishing gradients. This approach ensures that the network can be effectively trained even with increased depth, without requiring residuals at every layer.

By selectively adding residual connections, the overhead of summing residuals at every layer, which can introduce unnecessary complexity, is avoided. Instead, residuals were applied at key transition points where feature transformations occur, allowing the model to benefit from the stabilizing effects of residuals without overcomplicating the architecture.

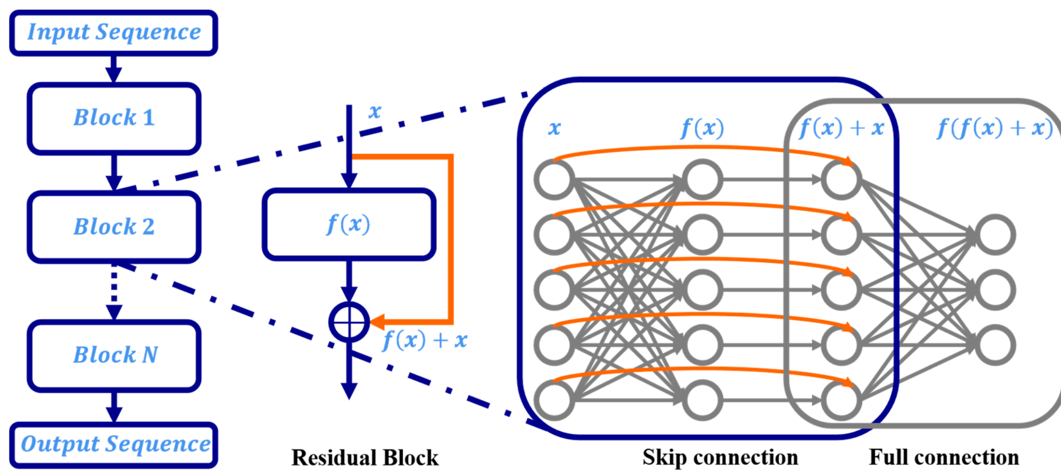


Figure 2. ResNet structure.

Each pair of hidden layers was treated as performing a specific level of feature transformation. With residuals introduced after every two layers, the model was enabled to capture and combine features from different levels more effectively, resulting in a richer representation of the input data. This selective approach allowed the model to scale and expand its learning capacity.

In shallow ANN structures, due to the presence of fewer hidden layers, the gradient vanishing problem is minimal. In this context, Sigmoid and tanh are appropriate activation functions. However, in deeper residual networks, using Sigmoid or tanh can lead to gradient diminishing across layers, rendering the early layers nearly untrainable. To address this issue, the Lintanh activation function was designed, combining the advantages of both linear and tanh functions. The Lintanh function is defined in Equation (1).

$$Lintanh(x) = \begin{cases} x, & if\ x \geq 0 \\ tanh(x), & if\ x < 0 \end{cases} \tag{1}$$

It maintains linearity for non-negative inputs, thereby preventing gradient vanishing, while providing non-linear transformation for negative inputs using tanh. Compared to Rectified Linear Unit (ReLU) and Leaky ReLU, which are commonly used in deep neural networks, Lintanh can capture more feature information. The smoothness and zero-centered output of tanh contribute to stable training and faster convergence. The first and second derivatives of Lintanh smoothly transition from negative to non-negative inputs, helping to avoid abrupt gradient changes and ensuring stable training. The Lintanh function is presented in Figure 3a, with the first and second derivative curves shown in Figures 3b and 3c, respectively.

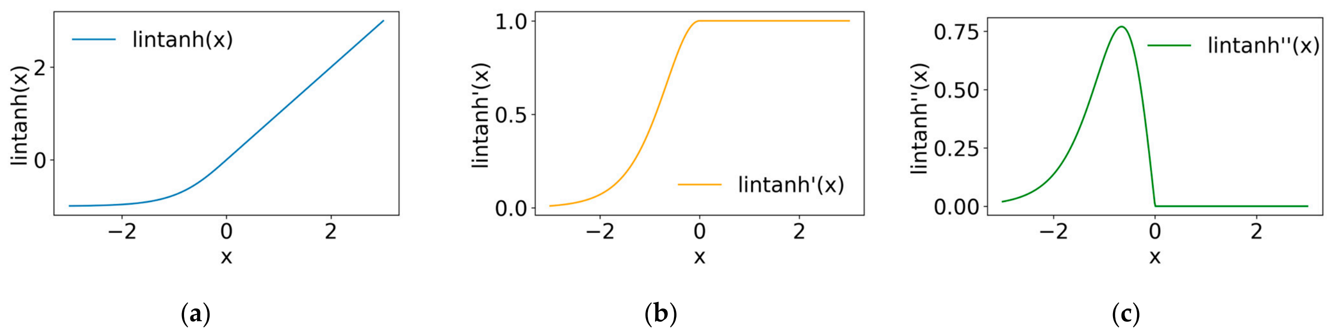


Figure 3. (**a**) Lintanh curves. (**b**) Lintanh's first derivatives curves. (**c**) Lintanh's second derivatives.

In contrast, the ReLU activation function, while popular in deep learning, has discontinuities in its first and second derivatives. Specifically, the first derivative of ReLU is not continuous at zero, transitioning abruptly from 0 to 1, and the second derivative is undefined at this point. The first derivative of ReLU can be expressed in Equation (2).

$$f'(x) = \begin{cases} 1, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases} \tag{2}$$

At $x = 0$, the derivative is not continuous, as the left-hand derivative is 0 and the right-hand derivative is 1. This lack of smoothness poses significant challenges in circuit simulation environments, particularly when using SPICE simulators. In these environments, device models, including surrogate models, often require the calculation of higher-order derivatives to accurately simulate the physical behavior of electronic devices. The matrix formulations used in SPICE simulators rely on the computation of multiple derivatives to ensure stable and accurate numerical results. The discontinuities in ReLU's derivatives can lead to numerical instability, making it less suitable for such simulations.

In contrast, Lintanh offers smooth first and second derivatives, making it a better fit for circuit modeling and simulation tasks that require accurate higher-order derivative calculations. The smoothness of Lintanh ensures that models built using this activation function can be seamlessly integrated into SPICE-like simulators, where the ability to compute continuous and stable derivatives is critical for simulating the dynamic behavior of devices under various conditions.

Considering the linear and non-linear aspects of the Lintanh activation function, Kaiming Uniform initialization was chosen to ensure stable output variance for each layer. Through uniform distribution, Kaiming Uniform initialization helps maintain good gradient flow in deep networks during early training.

Next, the AdaptiveSmoothL1 loss function was designed as written in Equation (3). When the actual data value is zero, Smooth L1 loss is applied, which is robust to outliers and suitable for handling zero targets. When the data value is non-zero, the absolute difference between the predicted and actual values is first computed and then the relative error is calculated. When the relative error exceeds 1, the loss function will speed up model convergence, enhancing learning in high-error scenarios. Conversely, if the relative error is below 1, the loss function will improve prediction accuracy, making the model more precise in low-error situations. It is seen that the proposed model can adjust flexibly to different situations, enhancing overall prediction performance.

$$AdaptiveSmoothL1 = \begin{cases} SmoothL1\left(y_{pred}, y_{true}\right), \ if\ y_{true} = 0 \\ \begin{cases} \frac{|y_{pred} - y_{true}|}{|y_{true}|} - 0.5, \ if\ \frac{|y_{pred} - y_{true}|}{|y_{true}|} > 1 \\ 0.5 \times \left(\frac{|y_{pred} - y_{true}|}{|y_{true}|}\right)^2, \ if\ \frac{|y_{pred} - y_{true}|}{|y_{true}|} \leq 1 \end{cases}, \ if\ y_{true} \neq 0 \end{cases} \tag{3}$$

Furthermore, when the batch size is set as 10, the initial learning rate of the Adam algorithm is set as 1×10^{-4}, and weight decay is set as 1×10^{-7}; the training will be stopped early to effectively prevent overfitting during surrogate model training.

Figure 4 shows the training and prediction flow of the ResNet model. The raw FinFET device data are preprocessed, including data enhancement, feature normalization, and logarithmic transformation, to form a dataset suitable for neural network training. The ResNet model was trained on the training set and the model performance was evaluated on the validation set at the end of each epoch to terminate the training early if overfitting occurred. The trained models are predicted on the test set and compared with the SPICE simulation values to evaluate the fitting and generalization ability of the models. In practical applications, the trained ResNet model can be used as a compact model to quickly and accurately predict the I–V and C–V characteristics of FinFET devices under any parameter combinations.

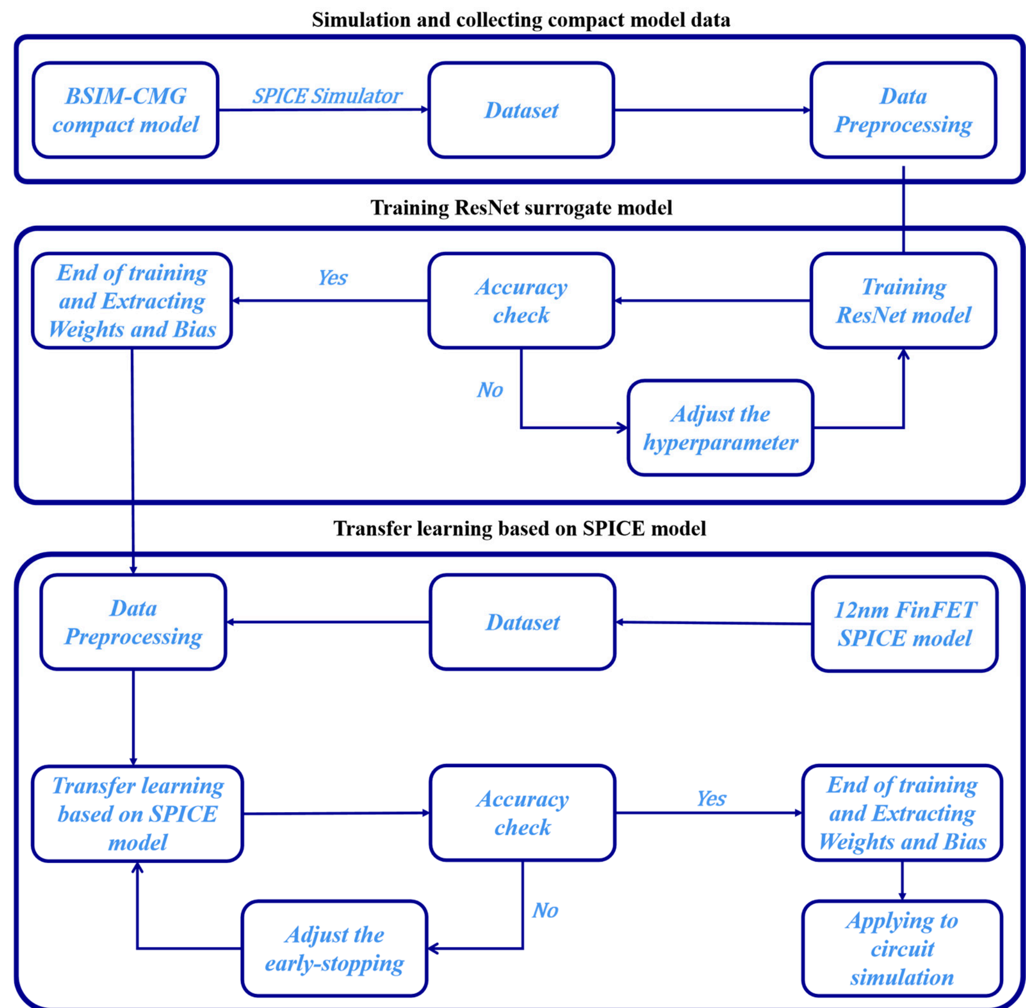
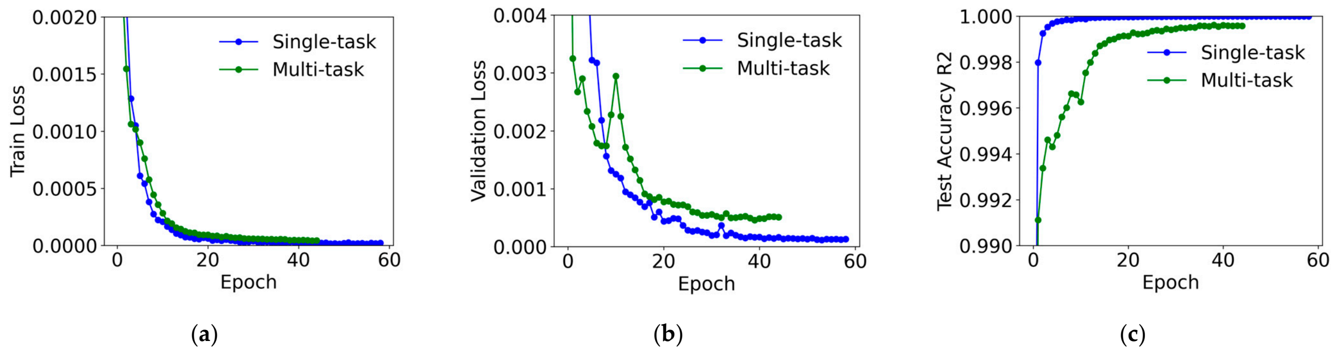


Figure 4. Training and prediction process of the ResNet model.

4. Results

Figure 5 shows the trends of the training loss (AdaptiveSmoothL1), validation loss (MAPE), and test set R^2 accuracy of the ResNet model in 12 nm FinFET single-task (I–V or C–V) and multi-task (I–V and C–V) modeling. During the training process, an early-stopping strategy was employed to prevent overfitting, with the training being terminated if the validation loss did not decrease for 6 consecutive epochs. This is consistent with the early-stopping strategy described earlier. As shown in Figure 5a, the optimal training loss of ResNet is 1.3962×10^{-5} (54th epoch) for single-task learning of I–V characteristics and 4.2310×10^{-5} (39th epoch) for multi-task learning of I–V and C–V characteristics. In both cases, the training loss converges quickly and stabilizes at a low level, indicating that ResNet effectively learns the nonlinear I–V and C–V characteristics of FinFET devices. The slightly higher training loss observed in multi-task learning compared to single-task learning is due to the increased modeling difficulty of fitting both I–V and C–V characteristics simultaneously. Figure 5b shows the variation of MAPE loss of ResNet on the validation set, where the best MAPE is 1.141×10^{-4} (54th epoch) for I–V single-task learning and 4.594×10^{-4} (40th epoch) for I–V and C–V multi-task learning. The trend of the validation loss is consistent with the training loss, reaching the optimum at the early stopping point and then slightly increasing, indicating the effectiveness of the early stopping strategy in preventing model overfitting. The validation MAPE of multi-task learning is slightly higher than that of single-task learning, but the overall control is still at a low level, as shown in Figure 5c. On the test set, the best R^2 for the I–V single-task is as high as 0.999992 (54th epoch), and the best R^2 for the I–V and C–V multi-task reaches 0.999579

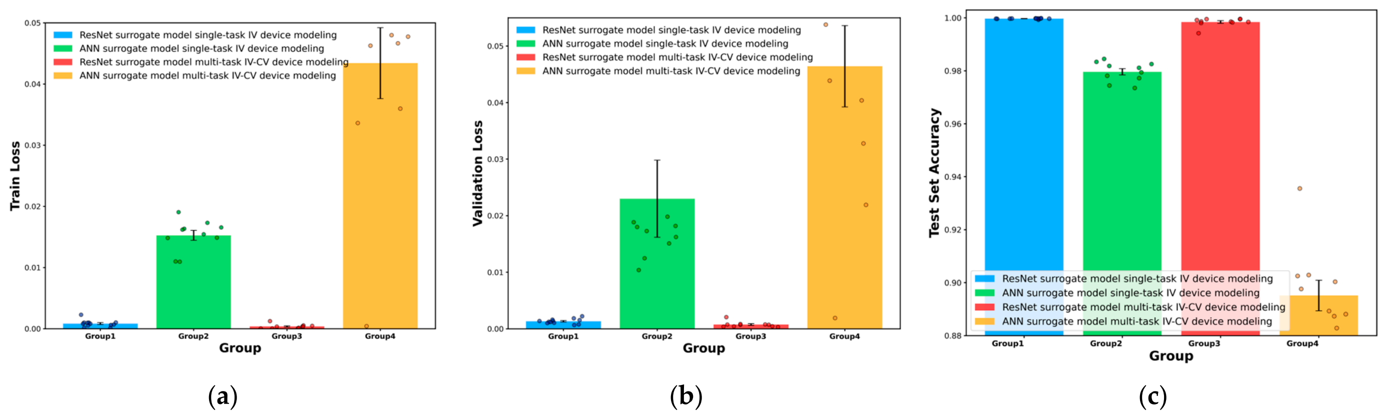
(45th epoch), and the goodness-of-fit of both of them is very close to 1. This indicates that ResNet can accurately reproduce the electrical characteristics of FinFET devices under different bias conditions, demonstrating its strong nonlinear mapping ability. It is noted that the R^2 of multi-task learning is slightly lower than that of single-task learning, but the difference between the two is small, which still meets the industry requirements for device modeling accuracy.



(**a**) (**b**) (**c**)

Figure 5. Comparison of ResNet model performance in 12 nm FinFET single-task (I–V) and multi-task (I–V and C–V) modeling with (**a**) training loss (AdaptiveSmoothL1); (**b**) validation loss (MAPE); and (**c**) test-set accuracy (R^2).

To further enhance the analysis, a comparative study between the ResNet and ANN models in both single-task and multi-task learning contexts is presented in Figure 6. Key performance indicators such as training loss, validation loss, and test set accuracy (R^2) are evaluated. As shown in Figure 6a, lower training loss (AdaptiveSmoothL1) is consistently achieved by the ResNet model compared to the ANN model. Similarly, in Figure 6b, better validation loss (MAPE) is observed for the ResNet model across both single-task and multi-task learning. Finally, in Figure 6c, higher test set accuracy (R^2) is reached by the ResNet model in comparison to the ANN model. The superior performance of ResNet is attributed to the use of residual connections, which are shown to mitigate gradient vanishing issues and improve the model's ability to capture complex nonlinear relationships in FinFET device modeling.



(**a**) (**b**) (**c**)

Figure 6. Comparison of ANN and ResNet model performance in 12 nm FinFET single-task (I–V) and multi-task (I–V and C–V) modeling for (**a**) training loss (AdaptiveSmoothL1), (**b**) validation loss (MAPE), and (**c**) test-set accuracy (R^2).

Figure 7 shows the I–V and C–V characteristic curves of some test set samples, comparing the ResNet prediction curves with the SPICE simulation curves. Figure 7a displays the Ids–Vgs characteristics, while Figure 7b–d present the Cgs–Vds, Cgd–Vds, and Cgb–Vds characteristics, respectively. In each sub-figure, the ResNet prediction curves are closely aligned with the SPICE simulation curves under all bias conditions, accurately capturing the I–V and C–V characteristics of FinFET devices and C–V characteristics of the FinFET device. It is demonstrated that the ResNet model has a strong nonlinear fitting ability and can accurately reproduce the trends of the Ids current and Cgs, Cgd, and Cgb capacitance of FinFET devices under different Vds and Vgs biases.

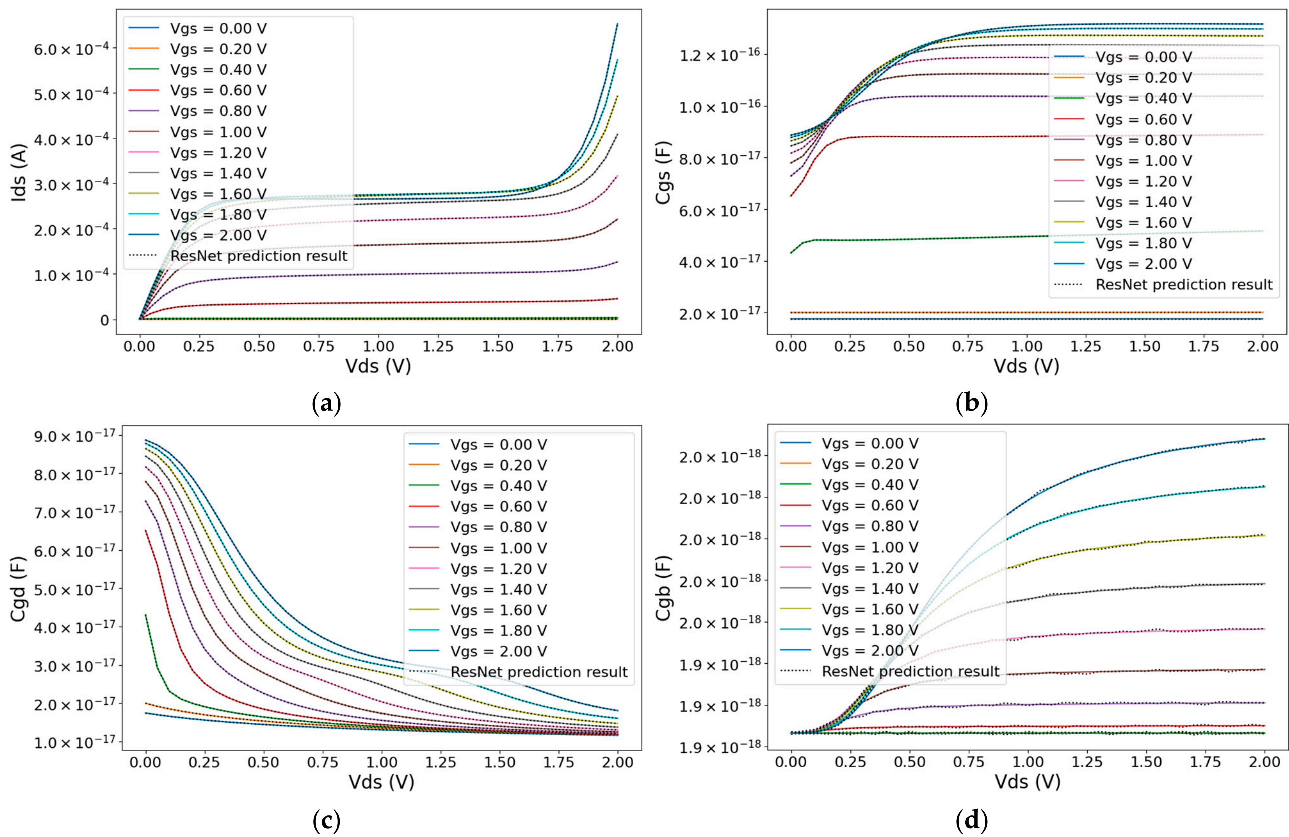


Figure 7. Comparison of ResNet Predictions and SPICE Simulations for FinFET I–V and C–V Characteristics. (**a**) Ids–Vds, (**b**) Cgs–Vds, (**c**) Cgd–Vds, and (**d**) Cgb–Vds.

Figure 8 presents the comparison of ResNet predicted values with SPICE simulated values on the FinFET device I–V and C–V characterization test set. There are four subplots: the Figure 8a Ids comparison, Figure 8b Cgs comparison, Figure 8c Cgd comparison, and Figure 8d Cgb comparison. The left side of each subfigure represents the ResNet prediction results and the right side is the SPICE simulation data. Scatter points are plotted on the VGS–VDS plane, and colors indicate the magnitude of the corresponding characteristics. In subfigures, it is shown that the color distributions of the ResNet predictions and SPICE simulations are highly consistent, confirming the high-precision fitting ability of the model.

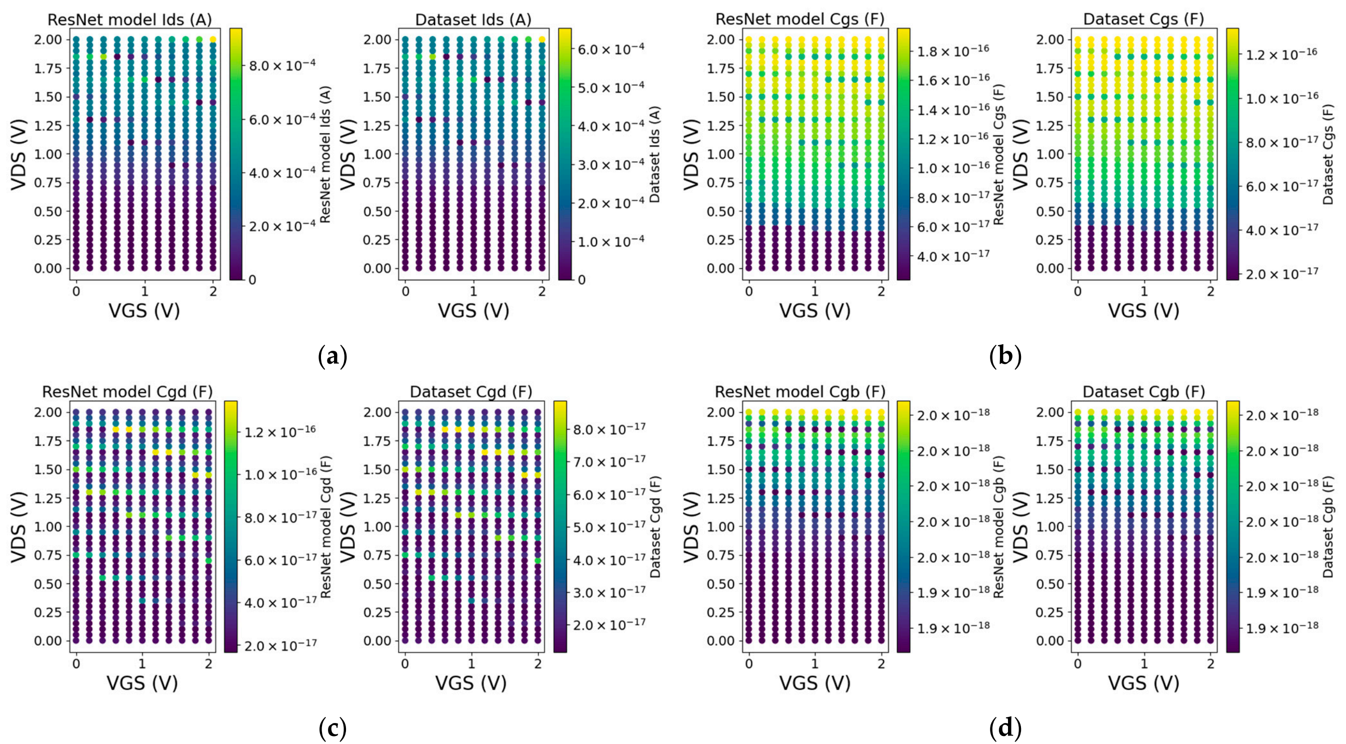


Figure 8. ResNet predictions vs. SPICE simulations: Scatter plots of I–V and C–V characteristics of FinFET devices. (**a**) Ids, (**b**) Cgs, (**c**) Cgd, and (**d**) Cgb.

Figure 9 is the performance of ResNet on generalization test sets, including single-task IV modeling (a), single-task IV transfer learning modeling (b), multi-task IV–CV modeling (c), and multi-task IV–CV transfer learning modeling (d). For single-task IV modeling, the average number of training rounds is 10, the average training loss is 8.36×10^{-4}, the average validation loss is 1.323×10^{-3}, and the average test R^2 is 0.99974. For single-task IV migration learning, the average number of training rounds is 1, the average training loss is 6.34×10^{-4}, the average validation loss is 1.265×10^{-3}, and the average test R^2 is 0.99975. The average test R^2 is 0.99973. It can be seen that utilizing migration learning can significantly reduce the number of rounds required for training, while maintaining a comparable level of loss and R^2 to that of ab initio training. For multi-task IV–CV modeling, the average number of training rounds is 16, the average training loss is 3.19×10^{-4}, the average validation loss is 6.81×10^{-4}, and the average test R^2 is 0.99442, whereas for multi-task IV–CV migration learning, the average number of training rounds is 5.6, the average training loss is 4.33×10^{-5}, the average validation loss is 2.22×10^{-4}, and the average test R^2 is 0.99973. It can be demonstrated that using migration learning can significantly reduce the number of training rounds required, while maintaining the same level of loss and R^2 as that of scratch training. The average training loss is 4.33×10^{-4}, the average validation loss is 2.22×10^{-4}, and the average test R^2 is as high as 0.99967. Multi-task migratory learning achieves better loss and R^2 performance than multi-task ab initio training while significantly shortening the training time.

Finally, a 13-stage RO circuit composed of FinFET devices was built, and Hspice simulations are performed based on a commercial SPICE model and ResNet model, respectively. The comparison of the characteristic curves of the two models is plotted in Figure 10. It is shown that the output voltage waveforms of the 13-stage RO simulated using the SPICE model and the ResNet model are highly consistent. Both models exhibit stable oscillation behavior, with the ResNet model accurately capturing the oscillation frequency and amplitude of the SPICE model.

Figure 9. Performance of ResNet model on FinFET generalization test set. (**a**) Single-task IV modeling, (**b**) single-task IV migration learning modeling, (**c**) multi-task IV-CV modeling, and (**d**) multi-task IV–CV migration learning modeling.

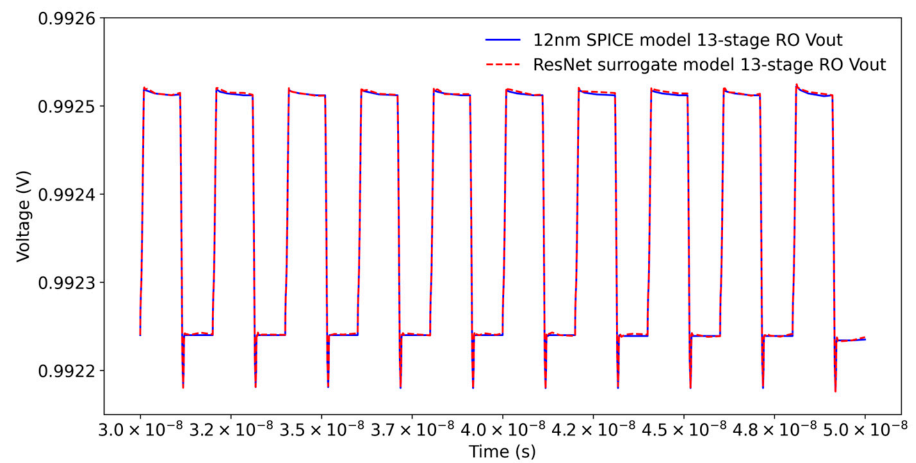


Figure 10. RO simulation curve SPICE model vs. ResNet model.

5. Discussions

In this study, a surrogate modeling method for FinFET devices based on ResNet was proposed, which incorporates a transfer learning strategy. Significant improvements in

modeling accuracy and efficiency were demonstrated by the ResNet model compared to traditional ANN models, particularly in the modeling of I–V and C–V characteristics of FinFET devices. The experimental results show that the nonlinear behavior of FinFET devices was accurately captured by the ResNet model, with strong generalization capabilities.

The BSIM-CMG compact model, which provides a comprehensive physical model for FinFET structures across a wide range of dimensions, was used as the foundation for this work. The performance of the compact model was successfully replicated by the ResNet-based surrogate model, suggesting its potential for generalization across different FinFET structures beyond the 12 nm technology node. This is a critical feature, as it indicates that future FinFET technologies, such as those at the 7 nm and smaller nodes, could be modeled by the ResNet model.

Similar to the ANN-based surrogate models discussed in previous studies [16–18], additional FinFET parameters, such as Vgs threshold voltage, drain-source on-resistance, and leakage currents, can also be predicted by the ResNet surrogate model. These parameters are critical for device evaluation, and, in future work, efforts will be made to integrate them into the ResNet model to make it a more versatile tool for device modeling.

Finally, while the use of transfer learning significantly accelerated the training of the SPICE surrogate model, further optimization of the transfer learning process remains possible. More advanced transfer learning techniques or fine-tuning strategies could potentially be explored to achieve further reductions in training time and improvements in accuracy.

6. Conclusions

In conclusion, a ResNet-based surrogate modeling method for FinFET devices was presented, with a transfer learning strategy incorporated to accelerate the modeling process. Superior accuracy and efficiency were demonstrated by the ResNet model compared to ANN models, particularly in the modeling of I–V and C–V characteristics. The BSIM-CMG compact model was used as the foundation for this work, ensuring that the underlying physical behavior of FinFET devices across various dimensions was accurately captured by the ResNet model.

Strong generalization capabilities were shown by the proposed ResNet surrogate model, making it suitable for application to FinFET structures beyond the 12 nm node. Future work is planned for validation at the 7 nm technology node. Additionally, the model is expected to be extended to predict other critical FinFET parameters, which would further enhance its utility as a device evaluation tool for DTCO.

With its demonstrated efficiency and accuracy, the proposed ResNet-based modeling method is considered a powerful approach to accelerating the design and optimization processes in advanced semiconductor technologies, offering a promising tool for future FinFET device evaluation and modeling.

Author Contributions: Conceptualization, Y.H.; methodology, Y.H.; software, Y.H.; validation, Y.H.; formal analysis, Y.H.; investigation, Y.H.; resources, B.L., Z.W. and W.L.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, B.L. and Z.W.; visualization, Y.H.; supervision, B.L. and Z.W.; project administration, B.L. and W.L.; funding acquisition, B.L., Z.W. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Guangdong S&T Programme, China (2022B0101180001).

Data Availability Statement: The datasets presented in this article are not readily available because PDK files are confidential and require declassification before use. Requests to access the datasets should be directed to the corresponding author.

Conflicts of Interest: Author Wenchao Liu was employed by the company Primarius Technologies Co. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Vetter, J.S.; DeBenedictis, E.P.; Conte, T.M. Architectures for the Post-Moore Era. *IEEE Micro* **2017**, *37*, 6–8. [CrossRef]
2. Liu, A.; Zhang, X.; Liu, Z.; Li, Y.; Peng, X.; Li, X.; Qin, Y.; Hu, C.; Qiu, Y.; Jiang, H. The roadmap of 2D materials and devices toward chips. *Nano-Micro Lett.* **2024**, *16*, 119. [CrossRef] [PubMed]
3. Li, Y.; Chen, D.; Tang, X.; Kong, L.; Li, L.; Deng, T. 3D-structured photodetectors based on 2D materials. *Appl. Phys. Lett.* **2024**, *124*, 13. [CrossRef]
4. Wang, X.; Liu, C.; Wei, Y.; Feng, S.; Sun, D.; Cheng, H. Three-dimensional transistors and integration based on low-dimensional materials for the post-Moore's law era. *Mater. Today* **2023**, *63*, 170–187. [CrossRef]
5. Cao, L.; Ren, C.; Wu, T. Recent advances in doped organic field-effect transistors: Mechanism, influencing factors, materials, and development directions. *J. Mater. Chem. C* **2023**, *11*, 3428–3447. [CrossRef]
6. Liao, M.; Sang, L.; Shimaoka, T.; Imura, M.; Koizumi, S.; Koide, Y. Energy-efficient metal–insulator–metal-semiconductor field-effect transistors based on 2D carrier gases. *Adv. Electron. Mater.* **2019**, *5*, 1800832. [CrossRef]
7. Nuytten, T.; Bogdanowicz, J.; Sergeant, S.; Fleischmann, C. Raman spectroscopy capabilities for advanced semiconductor technology devices. *Appl. Phys. Lett.* **2024**, *125*, 5. [CrossRef]
8. Zhu, J.; Palacios, T. Design–technology co-optimization for 2D electronics. *Nat. Electron.* **2023**, *6*, 803–804. [CrossRef]
9. Han, G.; Hao, Y. Design technology co-optimization towards sub-3 nm technology nodes. *J. Semicond.* **2021**, *42*, 020301. [CrossRef]
10. Gholipour, M.; Chen, Y.-Y.; Chen, D. Compact modeling to device-and circuit-level evaluation of flexible TMD field-effect transistors. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2017**, *37*, 820–831. [CrossRef]
11. Amer, S.; Hasan, M.S.; Adnan, M.M.; Rose, G.S. Spice modeling of insulator metal transition: Model of the critical temperature. *IEEE J. Electron Devices Soc.* **2018**, *7*, 18–25. [CrossRef]
12. Mohammed, M.U.; Nizam, A.; Ali, L.; Chowdhury, M.H. FinFET based SRAMs in Sub-10nm domain. *Microelectron. J.* **2021**, *114*, 105116. [CrossRef]
13. Valasa, S.; Ramakrishna, K.V.; Bhukya, S.; Narware, P.; Bheemudu, V.; Vadthiya, N. Performance investigation of FinFET structures: Unleashing multi-gate control through design and simulation at the 7 nm technology node for next-generation electronic devices. *ECS J. Solid State Sci. Technol.* **2023**, *12*, 113012. [CrossRef]
14. Dasgupta, A.; Parihar, S.S.; Kushwaha, P.; Agarwal, H.; Kao, M.-Y.; Salahuddin, S.; Chauhan, Y.S.; Hu, C. BSIM compact model of quantum confinement in advanced nanosheet FETs. *IEEE Trans. Electron Devices* **2020**, *67*, 730–737. [CrossRef]
15. Goel, R.; Wang, W.; Chauhan, Y.S. Improved modeling of flicker noise including velocity saturation effect in FinFETs and experimental validation. *Microelectron. J.* **2021**, *110*, 105020. [CrossRef]
16. Wang, J.; Kim, Y.-H.; Ryu, J.; Jeong, C.; Choi, W.; Kim, D. Artificial neural network-based compact modeling methodology for advanced transistors. *IEEE Trans. Electron Devices* **2021**, *68*, 1318–1325. [CrossRef]
17. Xia, K. A simple method to create corners for the lookup table-based MOSFET models through inputs and outputs mapping. *IEEE Trans. Electron Devices* **2021**, *68*, 1432–1438. [CrossRef]
18. Kao, M.-Y.; Kam, H.; Hu, C. Deep-learning-assisted physics-driven MOSFET current-voltage modeling. *IEEE Electron Device Lett.* **2022**, *43*, 974–977. [CrossRef]
19. Tung, C.-T.; Kao, M.-Y.; Hu, C. Neural network-based modeling with high accuracy and potential model speed. *IEEE Trans. Electron Devices* **2022**, *69*, 6476–6479. [CrossRef]
20. Jeong, H.; Woo, S.; Choi, J.; Cho, H.; Kim, Y.; Kong, J.-T.; Kim, S. Fast and expandable ANN-based compact model and parameter extraction for emerging transistors. *IEEE J. Electron Devices Soc.* **2023**, *11*, 153–160. [CrossRef]
21. Wei, J.; Wang, H.; Zhao, T.; Jiang, Y.-L.; Wan, J. A new compact MOSFET model based on artificial neural network with unique data preprocessing and sampling techniques. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **2022**, *42*, 1250–1254. [CrossRef]
22. Yang, Y.-S.; Li, Y.; Kola, S.R.R. A physical-based artificial neural networks compact modeling framework for emerging FETs. *IEEE Trans. Electron Devices* **2023**, *71*, 223–230. [CrossRef]
23. Yang, Z.; Gaidhane, A.D.; Anderson, K.; Workman, G.; Cao, Y. Graph-Based Compact Model (GCM) for Efficient Transistor Parameter Extraction: A Machine Learning Approach on 12 nm FinFETs. *IEEE Trans. Electron Devices* **2023**, *71*, 254–262. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.