Compact Modeling of N- and P-Type GAA NS FETs Using Physical-Based Artificial Neural Networks with Temperature Dependence

Yun Dei
Institue of Biomedical Engineering and
Parallel and Scientific Computing Laboratory
National Yang Ming Chiao Tung Univeristy
Hsinchu 300093, Taiwan.
ydei@mail.ymlab.org

Ya-Shu Yang
Institue of Communications Engineering and
Parallel and Scientific Computing Laboratory
National Yang Ming Chiao Tung Univeristy
Hsinchu 300093, Taiwan
ysyang@mail.ymlab.org

Yiming Li
Institue of Communications Engineering and
Parallel and Scientific Computing Laboratory
National Yang Ming Chiao Tung Univeristy
Hsinchu 300093, Taiwan
ymli@mail.ymlab.org

Abstract—We propose a compact model that utilizes physical-based artificial neural networks (ANNs) to model the effect of temperature on n- and p-type gate-all-around nanosheet FETs. Our compact model comprises two independent ANNs, where the first ANN is designed to output parameters related to temperature and the second ANN is utilized for the device physical parameters. All outputs of ANNs are integrated into a physical equation of drain current to form the entire compact model. Compared with the BSIM-CMG model in circuit simulations, our results are highly consistent in transfer characteristics and timing dynamics.

Keywords—Artificial neural networks, temperature dependence, physical-based compact modelling methodology, GAA NS MOSFETs.

I. INTRODUCTION

As semiconductor dimensions continue to shrink and structures become more complex, developing new CAD models and tools are increasingly difficult. The BSIM-CMG model requires a kind of methods to optimize over 600 equations [1]. Instead of using a multitude of physical equations, recent studies were reported with the physical-based single-ANN compact models [2], [3]. They can eliminate the need for laborious modifications of the physical equations for device models, but these studies have not considered the temperature effects yet. Temperature (T) is a crucial factor that affects the operation of chips [4]. High T increases carrier concentrations, leading to more leakage current. Thus, the temperature effect must be modeled in the ANN compact models.

In this work, we advance the physical-based ANN model for N- and P-type GAA nanosheet (NS) MOSFETs, where the temperature effect is included. It comprises two independent ANNs, where the first ANN is designed to model the temperature dependent parameters and the second ANN is utilized for the physical parameters of devices. Key parameters: T, gate oxide thickness (EOT), fin height (H_{FIN}), channel doping concentration (N_{BODY}), and gate workfunction (PHIG) are considered in both ANN models. Characteristic variations are discussed for T ranging from -25 to 105 °C.

II. MODEL ARCHITECTURE

The known ANN compact models were designed using a single ANN [5], [6]; differing from the existing works, we construct our model by using two separate networks: ANN1 and ANN2, as shown in Fig. 1. The output of our model, I_{ds}, is given by the following equation:

$$I_{ds} = \alpha \cdot \beta(T) \cdot \mu_0(T) \cdot C_{ox} \cdot \frac{W_{eff}}{L_{eff}} \cdot N_{FIN} \cdot V\left(V_{gs}, V_{ds}, V_{th}(T)\right), \quad (1)$$

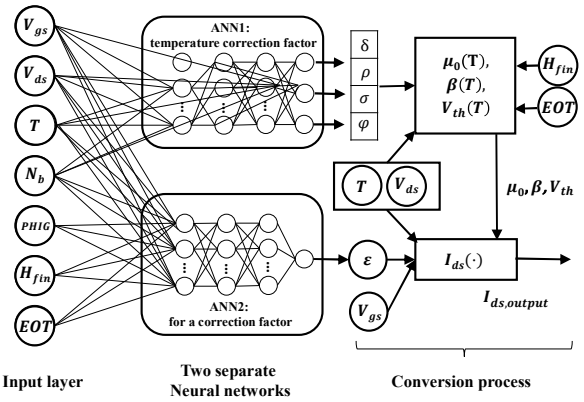

Fig. 1. An architecture of the ANN compact model includes a total of seven inputs, including V_{gs}, V_{ds}, temperature, and four physical parameters. Two separate hidden layers are used to enhance the influence of temperature on the model and improve its accuracy.

TABLE I. LIST OF THE HYPERPARAMETERS OF THE ANN1 AND ANN2.

ANN 1		ANN 2	
Parameters	***Values***	***Parameters***	***Values***
Hidden layers	4	Hidden layers	4
Neurons	10, 10, 10, 10	Neurons	10, 10, 10, 10
Epochs	1200	Epochs	1200
Learning rate	0.01	Learning rate	0.01
Scheduler	StepLR	Scheduler	StepLR
Activation function	tanh	Activation function	tanh
Loss function	MSE	Loss function	MSE
Input dimension	(16,4)	Input dimension	(16,7)
Output dimension	(16,3)	Output dimension	(16,1)
Optimizer	SGD	Optimizer	SGD

In (1), α and β(T) are temperature independent and dependent. $\mu_0(T)$ is the temperature-dependent mobility,

$$\beta(T) = 0.5 + 0.5 \times \tanh(\rho \times V_{ds}/T), \quad (2)$$

$$\mu_0(T) = \varphi/\sqrt{T} + \sigma \times \sqrt{T}. \quad (3)$$

C_{ox} is oxide capacitance, W_{eff} is width, L_{eff} is length, and N_{FIN} is the number of fins. $V(V_{gs}, V_{ds}, V_{th}(T))$ is a value calculated from V_{gs} , V_{ds} and $V_{th}(T)$.

$$V\left(V_{gs}, V_{ds}, V_{th}(T)\right) = V_{ds} \times \left(V_{gs} - V_{th}(T)/\sqrt{T}\right), \quad (4)$$

where

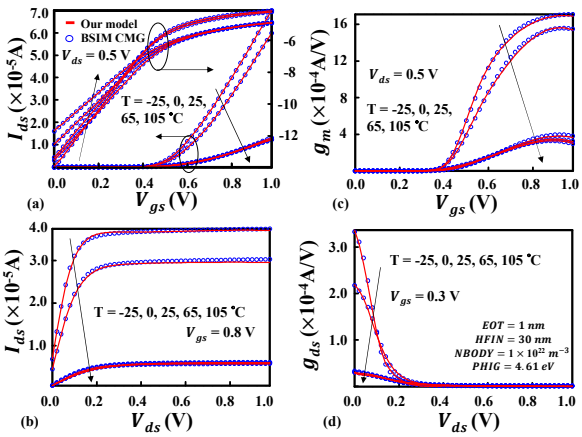

Fig. 2. The NMOS modeling results at different temperatures, with EOT = 1 nm, HFIN = 30 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV, are presented for (a) I_{ds}–V_{gs} at V_{ds} = 0.5 V, (b) I_{ds}–V_{ds} at V_{gs} = 0.8 V, (c) g_m–V_{gs} at V_{ds} = 0.5 V, and (d) g_{ds}–V_{ds} at V_{gs} = 0.8 V, respectively.

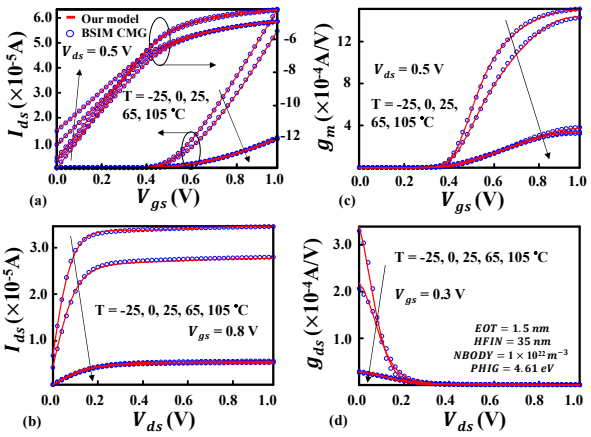

Fig. 3. The NMOS modeling results at different temperatures, with EOT = 1.5 nm, HFIN = 35 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV, are presented for (a) I_{ds}–V_{gs} at V_{ds} = 0.5 V, (b) I_{ds}–V_{ds} at V_{gs} = 0.8 V, (c) g_m–V_{gs} at V_{ds} = 0.5 V, and (d) g_{ds}–V_{ds} at V_{gs} = 0.8 V, respectively.

$$V_{th}(T) = \log\left(\delta/\sqrt{T} \times EOT \times Hfin\right). \quad (5)$$

The hyperparameters of ANN1 and ANN2 are listed in Tab. I. The ANN1 is responsible for producing a temperature-dependent function, while the ANN2 generates coefficients that are independent of temperature in (1). The four-layer ANN1 has four inputs: V_{os}, V_{ds}, T, as well as N_{BODY} and three outputs: δ, φ, ρ and σ. These four values serve as parameters for the functions β(T), μ_0(T) and V_{th}(T). By substituting T into (2), (3), and (5), the physically meaning parameters β, μ_0 and V_{th} in (1) can be determined.

We further construct the ANN2 with seven input nodes, four hidden layers, and one output. We utilize this output to obtain a correction term referred to as ε.

$$\varepsilon = 10^{x} = \alpha \cdot C_{ox} \cdot \frac{W_{eff}}{L_{eff}} \cdot N_{FIN}. \quad (6)$$

The input nodes comprise of V_{gs}, V_{ds}, N_{BODY}, T, EOT, H_{FIN}, and PHIG. From (6), x is the output of ANN2, where we can obtain α, C_{ox}, W_{eff}, L_{eff} and N_{FIN} in (1). Due to a nonlinear nature of I–V curves, (6) allows the proposed ANN compact

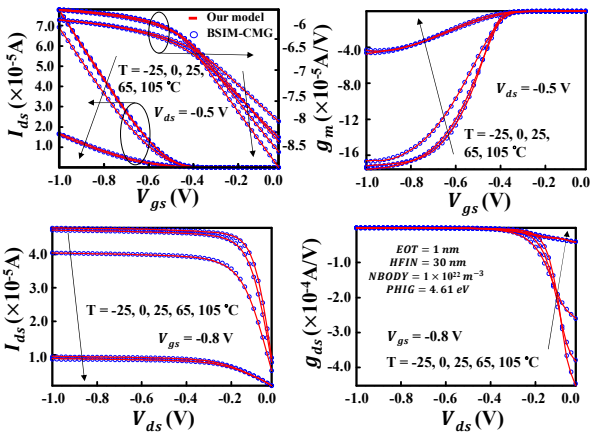

Fig. 4. The PMOS modeling results at different temperatures, with EOT = 1 nm, HFIN = 30 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV, are presented for (a) I_{ds}–V_{gs} at V_{ds} = -0.5 V, (b) I_{ds}–V_{ds} at V_{gs} = -0.8 V, (c) g_m–V_{gs} at V_{ds} = -0.5 V, and (d) g_{ds}–V_{ds} at V_{gs} = -0.8 V.

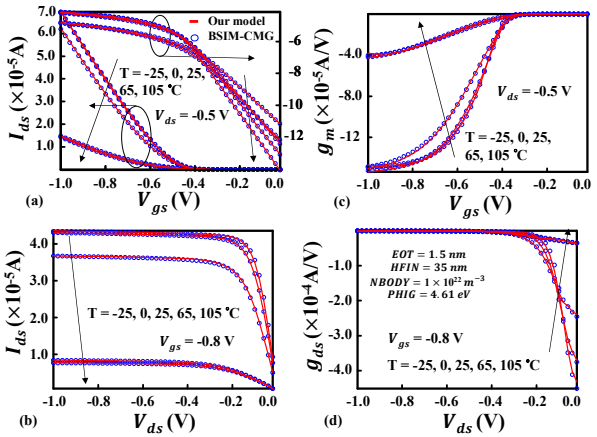

Fig. 5. The PMOS modeling results at different temperatures, with EOT = 1.5 nm, HFIN = 35 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV, are presented for (a) I_{ds}–V_{gs} at V_{ds} = -0.5 V, (b) I_{ds}–V_{ds} at V_{gs} = -0.8 V, (c) g_m–V_{gs} at V_{ds} = -0.5 V, and (d) g_{ds}–V_{ds} at V_{gs} = -0.8 V, respectively.

model to better capture the device characteristics. By combining the outputs from the ANNs, we can predict device characteristic under different biases and temperatures.

III. Results and Discussion

The ANN compact model is trained using 200,000 data points sourced from the BSIM-CMG model at temperatures of -25 °C, 0 °C, 25 °C, 65 °C, and 105 °C. The input V_{gs} and V_{ds} ranges are between 0 and 1 V in steps of 0.02 V. The loss function of the ANN compact model is given by:

$$\frac{1}{n}\left\{\begin{array}{c}\sum_{i=1}^{n}\left(I_{D,ANN}^{(i)} - I_{D,data}^{n(i)}\right) + \\ \sum_{i=1}^{n}\left[\log\left(I_{D,ANN}^{(i)}\right) - \log\left(I_{D,data}^{n(i)}\right)\right] \times \gamma\end{array}\right\}. \quad (7)$$

In order to improve the ANN compact model accuracy, we integrate a logarithmic function into the loss function. In our experiment, we conduct multiple tests and find that our compact model trained with γ of 0.008 achieved the highest model accuracy. Figs. 2-3 and Figs. 4-5 show the model results of N/PMOS devices of different sizes under different temperatures, respectively. The blue symbols are the data generated by the BSIM-CMG model, and the red lines

TABLE II. COMPARE THE V_{TH} AND SS EXTRACTED FROM PMOS DEVICES WITH DIFFERENT TESTING CASES USING THE ANN AND BSIM-CMG MODELS.

L_g = 30 nm, HFIN = 35 nm, EOT = 1.0 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV							L_g = 30 nm, HFIN = 30 nm, EOT = 1.0 nm, NBODY = 1×10^{23} m^{-3}, PHIG = 4.61 eV						
T (°C)	V_{th} (V)		SS (mV/dec)		Error (%)		T (°C)	V_{th} (V)		SS (mV/dec)		Error (%)	
	BSIM	*Ours*	*BSIM*	*Ours*	V_{th}	*SS*		*BSIM*	*Ours*	*BSIM*	*Ours*	V_{th}	*SS*
-25	-0.44	-0.44	50.25	50.28	0.451	0.060	-25	-0.451	-0.44	50.21	50.23	0.443	0.040
0	-0.43	-0.42	55.32	55.44	0.465	0.217	0	-0.441	-0.44	55.28	55.35	0.227	0.127
25	-0.43	-0.42	60.40	60.46	0.694	0.099	25	-0.443	-0.44	60.34	60.49	0.451	0.249
65	-0.52	-0.52	68.63	68.75	0.570	0.175	65	-0.541	-0.53	68.44	68.47	0.370	0.044
105	-0.49	-0.49	77.31	77.52	0.602	0.272	105	-0.514	-0.51	77.06	77.23	0.778	0.221

TABLE III. COMPARING ANN MODELS FOR DEVICE CHARACTERIZATION: INPUT/OUTPUT PARAMETERS AND ANALYSIS.

Model	**Ref. [6]**	**Ref. [3]**	**Ref. [10]**	**Ref. [9]**	**Ref. [7]**	**This Work**
Input parameters	V_{gs}, V_{ds}, T_{ox}, HFIN, N_{ch}, ϕ_M	V_{gs}, V_{ds}	V_{gs}, V_{ds}, L_C, T_{ox},	V_{gs}, V_{ds}, L, W	V_g, V_d, V_b, L, W, T	V_{gs}, V_{ds}, EOT, HFIN, NBODY, PHIG, T
Output parameters	I_{ds}	ε	I_{ds}	I_{ds}	I_{ds}	I_{ds}
Physical meaning	No	**Yes**	No	No	No	**Yes**
First derivative simulation result	No	**Yes**	No	**Yes**	No	**Yes**
Circuit simulaiton	No	No	No	**Yes**	**Yes**	**Yes**

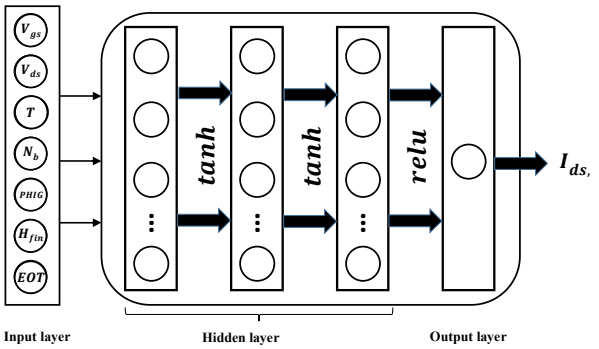

Fig. 6. An architecture of the conventional ANN compact model without physical meaning includes a total of seven inputs, including V_{gs}, V_{ds}, temperature, and four physical parameters.

represent the modeling results of the ANN compact model. The physical-based ANN model, even without incorporating the calculation of the first derivative in the loss function, accurately captures the characteristics of I_{ds}–V_{gs}, I_{ds}–V_{ds}, g_m, and g_{ds}. To further evaluate its performance, we compare our compact model with a non-physically meaningful ANN compact model [7].

Fig. 6 depicts an architectural of the non-physically meaningful compact model, while Fig. 7 illustrates the corresponding simulation results. Both models are trained on the same dataset and epochs. The non-physically meaningful compact model exhibited significant errors in I_{ds}–V_{gs}, I_{ds}–V_{ds}, g_m, and g_{ds}, with values of 1.57%, 1.66%, 8.44%, and 88.95%, respectively. In contrast, our simulated results show much lower errors of 1.33%, 1.01%, 1.88%, and 2.12%. This demonstrates that our ANN model with physical meaning provides more accurate values for the first-order derivative and produces smoother curves in simulations across different temperatures. Due to the use of the Rectified Linear Unit (ReLU) as the activation function in the last layer of Fig. 6, it offers faster training time for the model. However, it fails to provide accurate results when the data gradient is large. Furthermore, the activation of neurons in the last layer is not guaranteed when inputs are less than zero. Therefore, the hyperbolic tangent (tanh) function is a preferable choice. Additionally, the temperature significantly affects the I_{ds}. By incorporating various physical formulas related to V_{gs}, V_{ds}

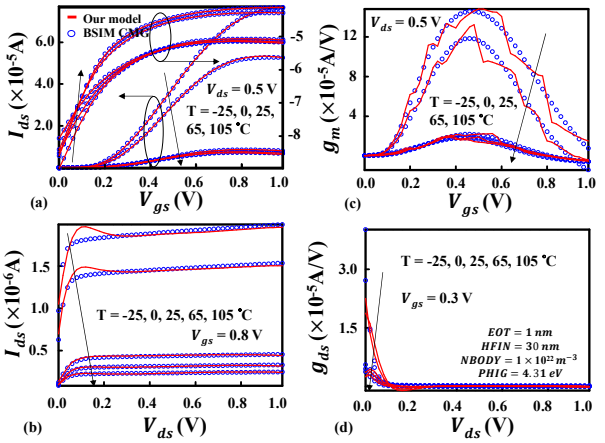

Fig. 7. The NMOS simulation results using nonphysiscal-meaning ANN and BSIM CMG models at different T, with EOT = 1 nm, HFIN = 30 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.31 eV, are presented for (a) I_{ds}-V_{gs} at V_{ds} = 0.5 V, (b) I_{ds}-V_{ds} at V_{gs} = 0.8 V, (c) g_m-V_{gs} at V_{ds} = 0.5 V, and (d) g_{ds}-V_{ds} at V_{gs} = 0.8 V, respectively.

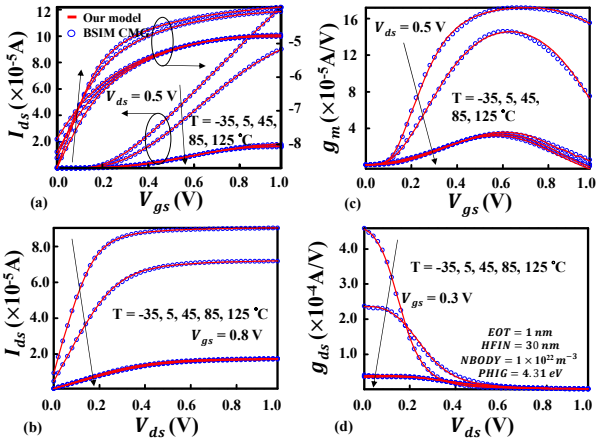

Fig. 8. The NMOS predictive simulation results using the physical-based ANN and BSIM CMG models at different T, with EOT = 1 nm, HFIN = 30 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.31 eV, are presented for (a) I_{ds}-V_{gs} at V_{ds} = 0.5 V, (b) I_{ds}-V_{ds} at V_{gs} = 0.8 V, (c) g_m-V_{gs} at V_{ds} = 0.5 V, and (d) g_{ds}-V_{ds} at V_{gs} = 0.8 V, respectively.

TABLE IV. RISE TIME (T_R) AND FALL TIME (T_F) OF A CMOS INVERTER.

L_g = 30 nm, HFIN = 30 nm, EOT = 1.0 nm, NBODY = 1×10^{22} m^{-3}, PHIG = 4.61 eV						
TEMP (°C)	t_r (µs)		t_f (µs)		Error (%)	
	BSIM	*Ours*	*BSIM*	*Ours*	t_r	t_f
-25	131.4	133.2	141.5	139.2	1.36	-1.6
0	147.3	153.5	158.1	160.2	4.20	1.32
25	191.5	193.0	213.7	218.5	0.78	2.24
65	318.0	324.2	319.9	322.3	1.94	0.75
105	510.0	506.1	478.4	491.2	-0.8	2.67

and T into the model, rather than treating them solely as inputs, the differentiation at different temperatures does not need to be explicitly included in the loss function, resulting in highly accurate results. Fig. 8 shows the I_{ds} predictions of an NMOS device at different T using the physical-based ANN compact model. Notably, even beyond the training temperature range of the ANN model, our model still yields highly accurate simulation results. The errors for I_{ds}–V_{gs}, I_{ds}–V_{ds}, g_m, and g_{ds}, are 0.93%, 0.8%, 1.21%, and 2.27%, respectively, indicating the model's ability to make accurate predictions outside the training range.

Based on the characteristic of PMOS devices, we do calculate the threshold voltage ($V_{th,nn}$) and subthreshold swing (SS_{nn}) in different T [8] and compared the results with V_{th} and SS obtained from the BSIM-CMG model in Tab. II. When T varies from -25 to 105°C, the errors are within 1%, except one at 2.4%. Tab. III compares our and recent models. Our model stands out with the highest number of input parameters, which indicates that the increased complexity in establishing an accurate model. Additionally, we observe that models with a physical meaning are capable of accurately simulating the first derivative values. While Ref. [9] lacks physical meaning, they incorporate the first derivative values in their loss function. The accuracy of the first derivative simulation directly impacts the accuracy of circuit simulation. Although Ref. [7] includes circuit simulation, the inaccurate first derivative simulation results in subpar circuit simulation performance. To evaluate the performance of the constructed ANN compact model in circuit simulation, we have converted the trained model into Verilog-A code and conducted circuit simulations using HSPICE® [9]. Fig. 9 shows the circuit simulation of the CMOS inverter. The blue symbols represent the data generated by the BSIM-CMG model, the red lines are the modeling results of the ANN compact model, and the black line are the input voltage of the CMOS inverter. Figs. 9(c)-(d) show the variations of the output voltage during rise and fall at different temperatures, respectively. It can be clearly observed that the time required for voltage rise and fall gradually increased from short to long as the temperature increased from low to high, consistent with the results shown by the BSIM-CMG model. Tab. IV indicates that the errors in the simulated rise time (t_r) and fall time (t_f) between the BSIM-CMG and our models are both less than 4.2% and an average is at 1.76%. Our ANN model results indicate that higher temperatures lead to an increase in carrier concentration and diffusion velocity in the GAA NS channel, resulting in higher channel resistance and longer t_r and t_f. Through the results of devices modeling and circuits composed of multiple devices, it shows that the proposed NN modeling work can successfully simulate at different temperatures with engineering acceptable accuracy.

IV. CONCLUSION

In summary, we have developed a physical-based compact model using two independent four-layer ANN architectures for GAA NS FETs from -25 to 105°C. Compared with recently reported works, seven parameters: V_{gs} ,V_{ds} , EOT, HFIN, NBODY, PHIG, T have been considered and validated in this work. Notably, the model was applied to various circuit simulations without any numerical divergence issues.

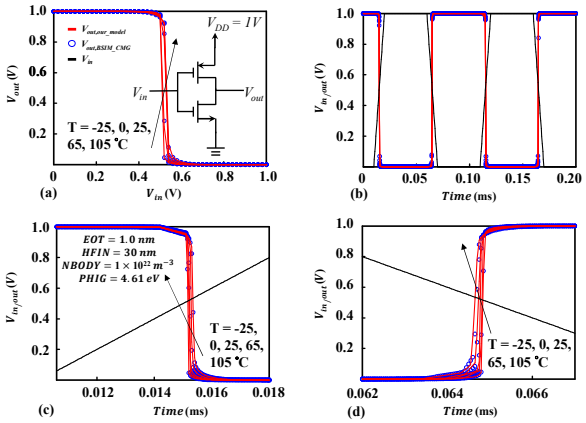

Fig. 9. Inverter simulation with our and the BSIM CMG models, where the devices are with EOT = 1 nm, HFIN = 30 nm, NBODY = 1×10^{22} m^{-3}, and PHIG = 4.61 eV. (a) Voltage transfer curves at different T, (b) Timing plots at different T. Zoom-in plots of timing variations, where (c) varies from 0.012 to 0.018 ms and (d) varies from 0.062 and 0.067 ms in different T.

ACKNOWLEDGMENT

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant MOST 111-2221-EA49-181, and NSCT 112-2218-E-006-009-MBK; in part by the "2022 Qualcomm Taiwan Research Program (NYCU)" under Grant NAT-487835 SOW.

REFERENCES

[1] J. P. Duarte et al., "BSIM-CMG: Standard FinFET compact model for advanced circuit design," in ESSCIRC, Austria, 2015, pp. 196-201.

[2] M. Li et al., "Physics-Inspired Neural Networks for Efficient Device Compact Modeling," IEEE JXCDC, vol. 2, pp. 44-49, Dec. 2016.

[3] M.-Y. Kao et al., "Deep-Learning-Assisted Physics-Driven MOSFET Current-Voltage Modeling," IEEE EDL, vol. 43, no. 6, pp. 974-977, June 2022.

[4] S. M. Sze, Y. Li and K. K. Ng, Physics of semiconductor devices, Wiley, 2021.

[5] L. Zhang, Chan, M. "Artificial neural network design for compact modeling of generic transistors," J. Comput. Electron., vol. 16, no. 3, pp. 825-832, Sep. 2017.

[6] S. Guglani et al., "Artificial Neural Network Surrogate Models for Efficient Design Space Exploration of 14-nm FinFETs," in DRC, OH, June 2022, 2pp.

[7] J. Wei et al., "A New Compact MOSFET Model Based on Artificial Neural Network With Unique Data Preprocessing and Sampling Techniques," IEEE T. Comput. Aid. D., vol. 42, no. 4, pp. 1250-1254, April 2023.

[8] J. Pathrose et al., "Temperature Sensor Front End in SOI CMOS Operating up to 250 °C," IEEE T. Circuits-II., vol. 61, no. 7, pp. 496-500, July 2014.

[9] Q. Yang et al., "Transistor Compact Model Based on Multigradient Neural Network and Its Application in SPICE Circuit Simulations for Gate-All-Around Si Cold Source FETs," IEEE Trans. Electron Dev., vol. 68, no. 9, pp. 4181-4188, Sep. 2021.

[10] K. Tamersit and F. Djeffal, "Fast and Accurate Simulation of Ultrascaled Carbon Nanotube Field-Effect Transistor Using ANN Sub-Modeling Technique, " in IEEE DTS, Tunisia, 2019, pp. 1-5.