A Machine Learning Approach for Optimization of Channel Geometry and Source/Drain Doping Profile of Stacked Nanosheet Transistors

Haoqing Xu, Weizhuo Gan, Lei Cao, Cheng Yang, Jiahao Wu, Mi Zhou, Hengze Qu, Shengli Zhang, Huaxiang Yin, *Senior Member, IEEE*, and Zhenhua Wu, *Member, IEEE*

***Abstract*— Complex nonlinear dependence of ultra-scaled transistor performance on its channel geometry and source/drain (S/D) doping profile bring obstacles in the advanced technology path-finding and optimization. A machine learning-based multi-objective optimization (MOO) workflow is proposed to optimize the sub-3-nm node gate-all-around (GAA) three-layer-stacked nanosheet transistors (NSFETs) accounting for the key performance knob of channel geometry and S/D doping profile. The artificial neural network (ANN) is trained to learn the compact current–voltage (*I–V*) relationship of NSFETs from 3-D technology computer-aided design (TCAD) simulation results. Based on the artificial neural network (ANN) model, MOO between threshold swing, on–off ratio, and ON-state current of NSFETs is performed with adaptive weighted sum theory. The proposed workflow efficiently suggests an optimized design window of channel geometry and doping profile of NSFETs. The proposed devices satisfy the 2025 International Roadmap for Devices and Systems (IRDSs) target in terms of electrical characteristics for digital circuits.**

***Index Terms*— Machine learning, multi-objective optimization (MOO), nanosheet, technology computer-aided design (TCAD) simulation.**

I. Introduction

THE critical dimension of electronic devices has always been scaling down in keeping with Moore's Law over the past decades. From planar to FinFETs, the technologies maintain the aggressive downscaling trend of devices [1]. However, when it comes to the 5-nm technology node, the ultra-scaled tri-gate devices perform limited gate control and suffer severe performance degradation due to the short channel effect (SCE). Thus, gate-all-around (GAA) devices, including nanowire field-effect transistors (NWFETs) and stacked NSFETs, are introduced to boost gate control capability [2]–[5] and proposed to be superior device structures for further scaling.

Compared with extremely scaled FinFETs, stacked NSFETs provide more device optimization options due to the design flexibility, among which the channel geometry [6], [7], channel material [8]–[10], and stress [11], [12] are intensively studied approaches for NSFET optimization. Carrier transport in NSFETs is highly sensitive to channel geometry, as nanosheet width has a unique impact on carrier mobility and transconductance according to the "narrow sheet effect" [6], and nanosheet thickness is critical for quantum confinement in 2-D channels, leading to a complex nonlinear dependency on device performance. Source/drain (S/D) doping profile is generally controlled by implantation and annealing in the process, which is crucial for effective channel length [13]. However, limited specifications for sub-3-nm technology node NSFETs were reported in previous studies [5], [14], [15], whereas detailed geometry and doping profile optimization schemes have not been investigated thoroughly for device optimization.

To obtain optimal design parameters of channel geometry and S/D doping profile of NSFET for high performance (HP), the nonlinear dependency of electrical characteristics on channel geometry and S/D doping profile should be accurately exploited. However, this modeling faces challenges as follows: 1) traditional design of experiment (DOE) based on technology computer-aided design (TCAD) is time-consuming with a large number of calculations; 2) quantum physics-based Standard FET models toward devices at scaling limit that could capture the nonlinearity between electrical characteristics and design parameters of devices requires much time and domain expertise to build [16]–[18]; and 3) the tradeoff between electrical characteristics should be carefully considered, especially when the optimization is performed in a multidimensional space composed of several parameters of channel geometry and doping profile of devices. A possible solution for challenges 1) and 2) is machine learning-based modeling. It has been demonstrated in previous work that the ANN has great

Manuscript received March 20, 2022; revised May 3, 2022; accepted May 12, 2022. Date of publication May 24, 2022; date of current version June 21, 2022. This work was supported in part by the International Partnership Program of the Chinese Academy of Sciences under Grant E1YH01X, MOST under Grant 2021YFA1200502, and in part by NSFC under Grant 91964202. The review of this article was arranged by Editor S.-M. Hong. *(Corresponding authors: Zhenhua Wu; Shengli Zhang.)*

Haoqing Xu, Weizhuo Gan, Lei Cao, Cheng Yang, Huaxiang Yin, and Zhenhua Wu are with the Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China, and also with the School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: wuzhenhua@ime.ac.cn).

Jiahao Wu is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China.

Mi Zhou is with The Information Technology Center, Open University of Sichuan, Chengdu 610072, China.

Hengze Qu and Shengli Zhang are with the College of Material Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhangslvip@njust.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2022.3175708.

Digital Object Identifier 10.1109/TED.2022.3175708

capability in capturing nonlinear relationships with high accuracy between device parameters and electrical characteristics, which shows a significant reduction of computation cost in predicting electrostatic potential [19], [20], current–voltage and capacitance–voltage relationship [21]–[23], threshold voltage [24], metal work function [25], and other figures of merit (FoMs) [26], [27]. The ANN is also employed in compact modeling for novel devices [28], [29] for advanced design-technology co-optimization (DTCO). For challenge 3), the performance-oriented optimization targeting multielectrical characteristics is considered as MOO, which has been applied in VLSI circuits design [30] and nanoscale device design [31]. However, most of these works focus on either the high prediction accuracy of device performance or great computation cost reduction of ANNs. The machine learning-based MOO for silicon nanoscale devices is seldom reported.

In this work, a compact I–V model of three-layer-stacked NSFETs is created by the ANN trained with samples generated from TCAD simulations, by which the relationship between I–V and device design parameters are captured. The ANN-based MOO between threshold swing, on–off ratio, and ON-state current is performed with adaptive weighted sum theory, and design windows are proposed for both p-type and n-type NSFETs. This article is arranged as follows. In Section II, the TCAD simulation method and machine learning method are introduced. In Section III, the MOO results are presented and discussed. Finally, the conclusion is made in Section IV.

II. Methodology

The overall workflow for NSFET optimization in this work is shown in Fig. 1, which is generally composed of three parts: TCAD simulation, machine learning, and MOO. The basic device model is generated from TCAD simulation and the machine learning captures the I–V relationship for NSFETs. The MOO based on the compact model from ANNs gives an optimal design of geometry and doping parameters of NSFETs. The following are the details of the three parts mentioned in Fig. 1.

A. TCAD Simulation

Three-dimensional three-layer-stacked NSFET model is built using commercial Global TCAD Solutions (GTS) Framework TCAD software [32], which is shown in Fig. 2(a). The doping profile of the NSFET in the channel and S/D extent region follows Gaussian distribution whose peak position is determined by the thickness of the overlap junction between S/D L_{ov} and channel and distribution width determined by the variance σ [see Fig. 2(b)]. The device model is calibrated to experimental results of 3-nm technology node NSFET referring to [5] as is shown in Fig. 3 where gate length L_g, spacer length L_{sp}, nanosheet height H, and nanosheet width W are 15, 5, 5, and 20 nm, respectively. The nanosheet is built with realistic rounding corners whose radius is 1 nm. The drain voltage is 0.65 V for nMOS (−0.65 V for pMOS) in the calibration, and gate voltage is swept from 0 to 0.7 V (0 to −0.7 V for pMOS).

The calibration is performed in the following steps [15]. First, the oxide thickness and metal gate work function are split into various values and simulated, from the results of which the threshold voltage V_{th} and sub-threshold swing (SS) of the device model could match with that of the reference [5]. Second, the carrier saturation velocity and contact resistance are adjusted to get a calibrated ON-state current I_{ON} in the saturation area. Although our calibrated NSFET model is slightly different from the reference, it is well-fit enough for TCAD simulation and electrical data analysis.

The transfer characteristics of various NSFETs are simulated using carrier mobility model MINIMOS-6 [33] which considers several carrier scattering models including phonon scattering, ionized impurity scattering [34], and surface roughness scattering [35]. To generate drain current I_d under different bias conditions for various NSFETs, the DOE is performed with different geometric and doping parameters as is listed in Table I. Two channel geometric parameters describing the shape of the nanosheet, W and H, and two doping profile parameters, L_{ov} and σ, are split into various values with specific steps. The diffusion of dopant from the S/D region into channel is described using Gaussian distribution

$$g(x; L_{ov}, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(x \pm \left(\frac{L_g}{2} + L_{sp} - L_{ov}\right)\right)^2}{2\sigma^2} \right) \tag{1}$$

where the peak position is in S/D extent regions with distant L_{ov} from the source(drain)/spacer interface. Both p-type and n-type NSFETs are simulated where V_g is swept from +0.2 to −0.8 V with −0.1-V step for pMOS (−0.2 to +0.8 V with +0.1-V step for nMOS), and $V_{d\text{lin}}$ and $V_{d\text{sat}}$ are set to be −0.65 V and −0.05 for pMOS (+0.65 and +0.05 V for nMOS). All simulation tasks are performed, and the drain current of various NSFETs under different bias conditions are generated. Then the dataset is cleaned for training ANNs in the next steps.

B. Machine Learning Methodology

Data normalization is an important step before neural network training, which could reduce the training time and improve the accuracy of prediction. In this work, 31 680 (31 680 = 5*4*9*4*2*11*2) samples are generated from TCAD and each sample has seven features (H, W, L_{ov}, σ, MOSType, V_g, and V_d) and one label (I_d) as listed in Table I. The normalization of the TCAD data is generally composed of the following two steps.

1) *Data preprocessing:* The feature MOSType is set to be −1 for p-type devices and +1 for n-type devices. The label I_d takes the negative logarithm of the absolute values of the original drain current, of which the distribution is much more suitable for normalization.
2) *Standard scaling:* The following equation represents the standard scaling:

$$x_{\text{norm}} = \frac{x - \mu}{s} \tag{2}$$

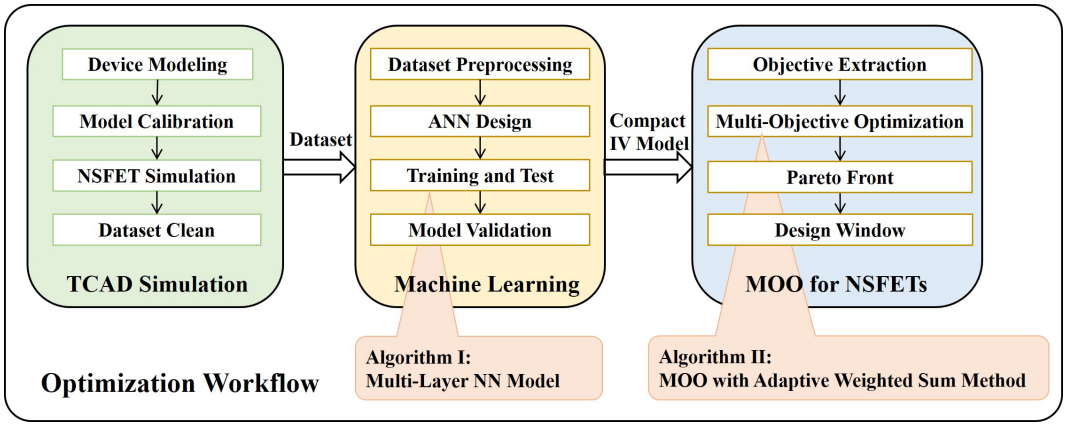

Fig. 1. Schematic of overall workflow for NSFET optimization.

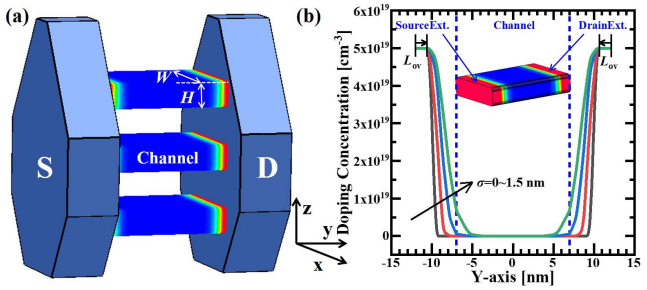

Fig. 2. (a) Schematic of the three-layer-stacked NSFET model used in this work and (b) doping profile following various Gaussian distribution in NSFETs.

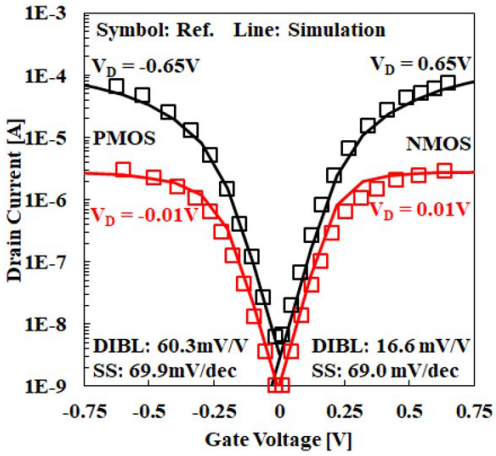

Fig. 3. NSFET calibration results of I_d–V_g curve with [5].

where μ is the mean value and s is the standard deviation of all samples, which is applied to the whole dataset. Centering and scaling happen independently on each feature or label by computing the relevant statistics on the samples in the dataset. After the normalization, all the samples are split into 60%, 20%, and 20% for train, validation, and test sets, respectively.

Fig. 4(a) shows the basic structure of the multilayer neural network used in this work, which is composed of a single input layer (a row vector X with the size of 7×1) regarding NSFET design parameters and bias conditions, a single output layer (a row vector Y with the size of 1×1) for predicted drain current, and two hidden layers with 15 nodes for each of them. For accurate training of the model, the sigmoid function

$$f = \frac{1}{1 + e^{-x}} \tag{3}$$

is used as the activation to capture the nonlinearity between the input and output. The neural network is implemented in Python with the machine learning framework PyTorch [36].

The network could be written in an analytical form as

$$I_d = W_3 \cdot f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) + b_3 \tag{4}$$

$$X = \left(H, W, L_{\text{ov}}, \sigma, \text{MOSType}, V_d, V_g\right) \tag{5}$$

where weight matrices W and bias vectors b are annotated by the given layer number i. The weight matrices and bias vectors are updated in the training using backpropagation (BP)

TABLE I
DESIGN PARAMETERS AND BIAS CONDITIONS FOR NSFETS

Parameter	Calibration	Range (min, max)	TCAD Step	MOO Step
H [nm]	5	(4, 6)	0.5	0.1
W [nm]	20	(15, 30)	5	1
L_{ov} [nm]	1	(0, 4)	0.5	0.1
σ [nm]	1.5	(0, 1.5)	0.5	0.1
MOSType	P, N	P, N	-	-
Vg [V]	-	(-0.2, 0.8)	0.1	0.02
Vd [V]	0.01, 0.65	0.05, 0.65	-	-

algorithm [37]. The performance in training NN is evaluated by the loss function mean squared error (MSE) defined by

$$\text{MSE} = \frac{1}{n}\sum_{i=0}^{n-1}\left(I_{d,i} - \hat{I}_{d,i}\right)^2 \tag{6}$$

where $I_{d,i}$ is the predicted drain current by the ANN and $\hat{I}_{d,i}$ is the drain current in the test set. The loss function is optimized by the Adam optimizer with the learning rate at 5.0×10^{-3} [38]. The MSE of both train set and validation set keeps decreasing as a function of epochs during the training process as is shown

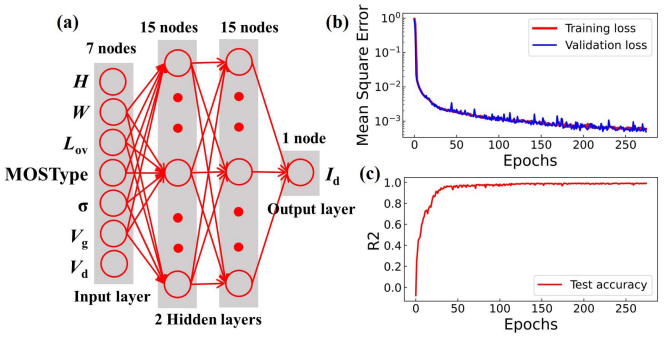

Fig. 4. (a) Schematic of the multilayer ANN structure trained in this work. (b) MSE works as loss function on the train set and validation set in the NN training and (c) prediction accuracy R^2 on test set.

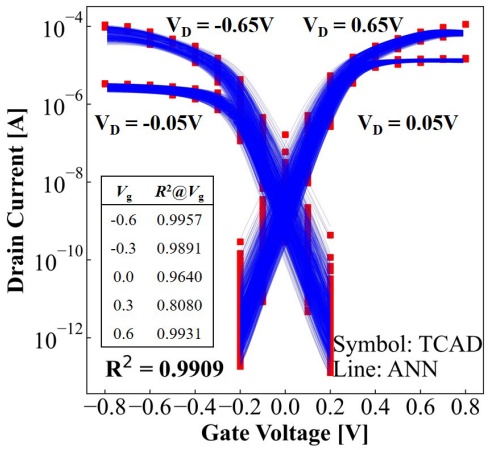

Fig. 5. Predicted and actual drain current versus gate voltage on the test set, showing the trained ANN performs well in capturing the current–voltage relationship.

in Fig. 4(b). The NN is determined when the error rate (MSE) on validation set $\leq 0.5\%$ at the 275 epoch.

Algorithm I Multilayer NN Model

Step 1: Normalize data generated from TCAD simulation
Step 2: Split dataset into train, test and validation sets
Step 3: Initialize multi-layer NN parameters
Step 4: **for** epoch in range(Epochs):
$I_{d,pred} = W_3 \cdot f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) + b_3$
CalcuSlate loss function
if MSE on validation set $\leq 0.5\%$:
Break
else:
Update weights and bias using BP
Go **Step 4**
Step 5: Test ANN performance on test set

Once the neural network is determined, the accuracy of drain current I_d prediction is analyzed on the test set by the coefficient of determination R^2 defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(I_{d,i} - \hat{I}_{d,i}\right)^2}{\sum_{i=1}^{n}\left(I_{d,i} - \bar{I}_{d,i}\right)^2} \tag{7}$$

where $\bar{I}_{d,i}$ is the mean value of drain current in the test set. The accuracy is calculated using the inverse normalization value of output of the NN. The overall R^2 on test set increases in the training process [see Fig. 4(c)] and reaches 0.9909 at the end of training. The predicted drain current is compared with actual drain current which matches well as shown in Fig. 5. The well-predicted drain current of various NSFETs with different geometric and doping parameters indicate that the NN has extracted the nonlinearity relationship between the input and outputs with high accuracy in the form of (4).

C. Multiobjective Optimization

SS, on–off ratio (I_{ON}/I_{OFF}), and normalized ON-state current (I_{ON}) are three FoMs of NSFETs to be highly concerned in device design and optimization. These three performance indicators are extracted from the trained NN model [see (4)] as follows. The SS of the device is defined as $SS = dV_g/d\log(I_d)$, and measured at $V_g = V_{th} \pm 50$ mV (positive for pMOS and negative for nMOS), where V_{th} is V_g at $I_d = 10$ nA·W/L under $V_d = \pm 0.65$ V. The on–off ratio is extracted from the transfer curves by $I_{ON}/I_{OFF} = I_{d,V_g = \pm 0.65V}/I_{d,V_g = 0V}$. The ON-state current is drain current under $V_g = V_{dd} = \pm 0.65$ V normalized by effective width.

These extracted FoMs are important indicators for NSFET performance. Generally, lower SS, higher on–off ratio, and higher ON-state current are preferred for device optimization. These FoMs are nonlinearly coupled and the optimization of FoMs should be considered an MOO problem. Therefore, multiobjective formulations are required for NSFET optimization.

One of the most widely used theories for the MOO problem is the weighted sum method that adds all target objectives with weights and converts a multiobjective problem into a maximum or minimum problem of a scalar [39]. And by changing the weight of objectives, various Pareto optimal solutions could be obtained [40]. However, as the usual weighted sum method produces poorly distributed solutions and concave parts cannot be detected, the adaptive weighted sum method is implemented for the MOO problem [41]. In this work, as there is a tradeoff between SS and I_{ON} as well as on–off ratio and I_{ON}, the utility functions are designed as

$$U_1 = \lambda_1 \frac{SS}{sf_{SS}} + (1 - \lambda_1)\frac{(-1) * I_{ON}}{sf_{I_{ON}}} \tag{8}$$

$$U_2 = \lambda_2 \frac{\log(I_{ON}/I_{OFF})}{sf_{\text{on-off-ratio}}} + (1 - \lambda_2)\frac{(-1) * I_{ON}}{sf_{I_{ON}}} \tag{9}$$

where λ_1 and λ_2 are weighting factors in U_1 and U_2, respectively, and sf_{SS}, $sf_{I_{ON}}$, and $sf_{\text{on-off-ratio}}$ are normalization scaling factors for the corresponding objectives. The MOO problem could be defined as

$$\begin{aligned} &\min\ U_1 \text{ (or } U_2) \\ &\text{s.t. } SS \leq 80\,\text{mV/dec (or } \log(I_{ON}/I_{OFF}) \leq 4.5) \\ &\quad \text{and } \lambda \in [0, 1]. \end{aligned} \tag{10}$$

By searching for the minimum of U_1 and U_2 in objective space with various weights of objectives, Pareto optimal

solutions are obtained, which offers a group of optimized design parameters for NSFET channel geometry and S/D doping profile. The algorithm of MOO with adaptive weight sum is shown in Algorithm II [41]. The prescribed constant are $n_{\text{initial}} = 20$, $\epsilon = 0.03$, $C = 2$, and $\delta_J = 0.06$.

Algorithm II MOO with Adaptive Weighted Sum Method

Step 1: Extract objectives from ANN model
Step 2: Identify objective space range
Step 3: Normalize the objective functions by Min-Max Scaling and define utility function U_1 and U_2
Step 4: Perform MOO using the usual weighted sum method with a small number of divisions n_{initial}
Step 5: Delete the nearly overlapping solutions if the distance among solutions is less than the prescribed distance ϵ
Step 6: Determine the number of further refinement in each of region
$n_i = round(C(l_i/l_{avg}))$
where l_i is the length of ith segment, l_{avg} is the mean value of length of all the segment, and C is a prescribed constant
Step 7: **If** $n_i \leq 1$:
End
Else:
Go Step 8
Step 8: Determine the offset distances from two end points of each segment
$\delta_1 = \delta_J \cos\theta$, $\delta_2 = \delta_J \sin\theta$, $\theta = \tan^{-1}(\text{-k})$
where k is the slope of the two end points, and δ_J is a prescribed distance
Step 9: Implement new inequality constraints and do sub-optimization using weighted sum method in each of the sub-regions
Step 10: Delete the nearly overlapping solutions if the distance among solutions is less than the prescribed distance ϵ
Step 11: **If** all $l_i \leq \delta_J$:
End
Else:
Go Step 6

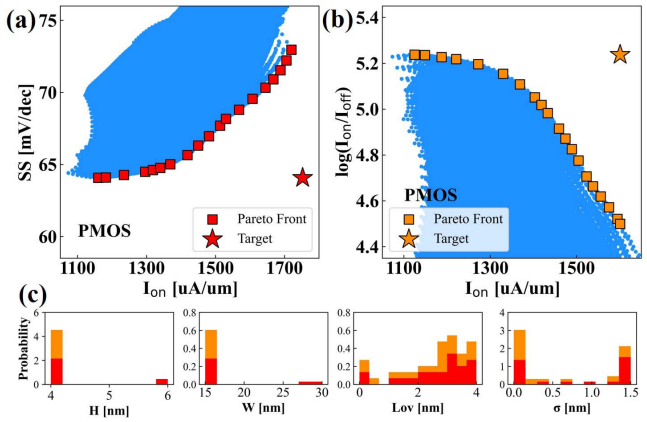

Fig. 6. Pareto front of (a) SS versus I_{ON}, (b) log(I_{ON}/I_{OFF}) versus I_{ON} as two objectives for p-type NSFETs, and (c) their Pareto optimal solution distribution.

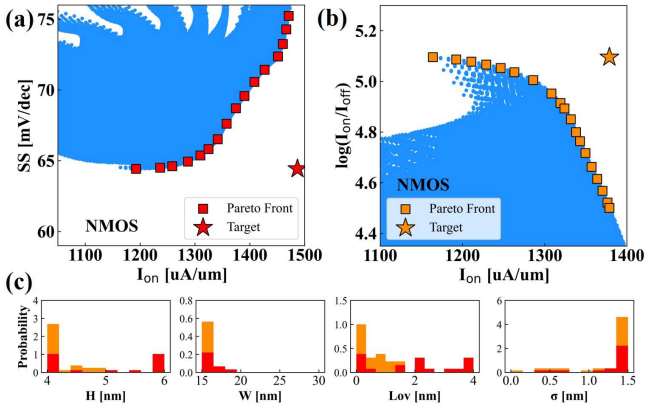

Fig. 7. Pareto front of (a) SS versus I_{ON}, (b) log(I_{ON}/I_{OFF}) versus I_{ON} as two objectives for n-type NSFETs, and (c) their Pareto optimal solution distribution.

III. Results and Discussions

A. Design Windows

First, p-type NSFETs are discussed. SS and I_{ON} are considered as a pair of two objectives for optimization as a tradeoff between them is required, of which the utility function is U_1 in (8). The objective domain is determined by the parameters range listed in Table I, where smaller steps of design parameters of the devices are used as inputs to calculate their corresponding I–V curve using the NN, and SS and I_{ON} are extracted from the I–V curves, which is shown as the blue area in Fig. 6(a). The solutions are distributed along Pareto Front of SS versus I_{ON}, which are shown as red squares in Fig. 6(a). The target shown as a small star at the right bottom corner, showing the preferred direction of optimization, lower SS, and higher I_{ON}. The distribution of four optimal design parameters for p-type NSFETs at Pareto Front is shown in Fig. 6(c) with red bars. Similarly, log(I_{ON}/I_{OFF}) and I_{ON} are paired and their MOO results under utility function U_2 are shown in Fig. 6(b) with orange squares. The distribution of its Pareto optimal solution is also shown in Fig. 6(c) with orange bars. From Fig.6(c), it could be clearly seen that ultra-thin layer at $H = 4$ nm, sheet width $W = 15$ nm, $L_{ov} = 3$ nm, and σ at zero are preferred for pMOS optimization. The reason is that the ultra-thin sheet is demonstrated to have strong gate control. W at lower limit is required as narrow sheet provides higher normalized I_{ON}. $L_{ov} = 3$ nm is a result of a tradeoff between gate control and ON-current as small L_{ov} leads to longer effective channel length but higher channel parasitic resistance and vice versa. σ at zero also enables the device to have a longer effective channel length and suffer less SCE.

N-type NSFET optimization results are shown in Fig. 7. Similarly, the MOO results for U_1 and U_2 are distributed along Pareto front in Fig. 7(a) and (b), respectively, and distribution of the Pareto optimal solutions are shown in Fig. 7(a) and (c), which suggests a thin and narrow nanosheet for the best tradeoff which shows the same preference to that of pMOS. However, different from pMOS optimization in Fig. 6(c), a wider Gaussian distribution of S/D doping profile parameter $\sigma = 1.5$ nm and a small L_{ov} are suggested [see Fig.7(c)], by which the wide distribution of S/D doping compensates for the loss of ON-state current due to increase in parasitic resistance by small L_{ov}.

NSFET optimal channel geometry and doping profile design windows could be extracted from Fig. 6(c) to 7(c) for both p-type and n-type devices, from which the value of design parameters distributed with max probability is chosen from the optimal design as shown in Table II. The performance of the device with optimal design is verified in TCAD and the digital

TABLE II
OPTIMAL DESIGN PARAMETERS FOR NSFETS

Mostype	H [nm]	W [nm]	L_{ov} [nm]	σ [nm]
PMOS	4.0	15.0	3.0	0.0
NMOS	4.0	15.0	0.0	1.5

TABLE III
OPTIMAL DESIGN PARAMETERS FOR NSFETS WITH FIXED σ

Mostype	H [nm]	W [nm]	L_{ov} [nm]	Fixed σ [nm]
PMOS	4.0	15.0	4.0	0.5
NMOS	4.0	15.0	3.0	
PMOS	4.0	15.0	4.0	1.0
NMOS	4.0	15.0	2.0	
PMOS	4.0	15.0	3.0	1.5
NMOS	4.0	15.0	1.0	

TABLE IV
PERFORMANCE COMPARISON OF OPTIMIZED NSFETS WITH INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEM (IRDS) TARGETS

	PMOS	NMOS	2025 Target	2028 Target
L_g [nm]	14	14	14	12
V_{dd} [V]	-0.65	0.65	0.65	0.65
SS [mV/dec]	64.8	61.6	72	75
DIBL [mV/V]	28.4	30	-	-
I_{on}/I_{off}	9.86×10^4	9.70×10^4	8.73×10^4	9.24×10^4
I_{on} [uA/um]	1429	1155	873	924
V_t [mV]	167	148	212	226

circuit in the following part. Accounting for a realistic process, the idea of the abrupt junction with $\sigma = 0$ can be hardly achieved. Thus, a group of specific σ from 0.5 to 1.5 nm are set for device optimization, where MOO problems with fixed σ are solved by Algorithm II. The resulting optimal design parameters are shown in Table III. It could be seen that under all fixed σ values, a thin and narrow nanosheet is preferred in device optimization so as to obtain good gate control for low SS and high on–off ratio. L_{ov} decrease as σ increases, and the balance between them shows that the effective channel length is kept for a balance between high I_{ON} and less SCE.

B. Discussion

As the critical dimension of devices scaling down beyond the 3-nm node, the performance of devices is expected as listed in International Roadmap for Devices and System (IRDS). The optimized NSFET models in Table II are simulated by TCAD and their performance is compared with that in IRDS targeting "2.1 nm" (2025) and "1.5 nm" (2028) technology node as shown in Table IV. To further assess the performance of the optimized NSFETs, the NSFETs performance of 32-bit ALU circuits is evaluated through software beyond-CMOS benchmark (BCB) [42]. Fig. 8 shows the switching energy and delay of proposed sub-3-nm NSFETs 32-bit arithmetic logic unit (ALU), of which the performance reaches (Energy/Daley = 145 fJ/1546 ps).

As we can see from Table IV and Fig. 8, the optimal NSFETs outperform devices in 2025 and 2028 IRDS target in terms of SS, drain induced barrier lowering (DIBL), on–off ratio, and normalized I_{ON} as well as that for 32-bit ALU performance. Both SS and DIBL of the optimized NSFETs are well tuned due to optimal doping profile leading to the longer effective channel length. Both threshold voltage of n-type and p-type NSFETs matches well. However, to compromise for higher I_{ON}, n-type NSFET on–off ratio degrades slightly compared with that of p-FET. A possible approach for further optimization of n-type NSFETs performance is introducing stress in the channel for career mobility enhancement. In addition, C–V characteristics are also important for the performance power benchmark. In principle, the proposed ANN model and MOO scheme are also capable to evaluate and optimize C–V characteristics. It will be addressed in further studies.

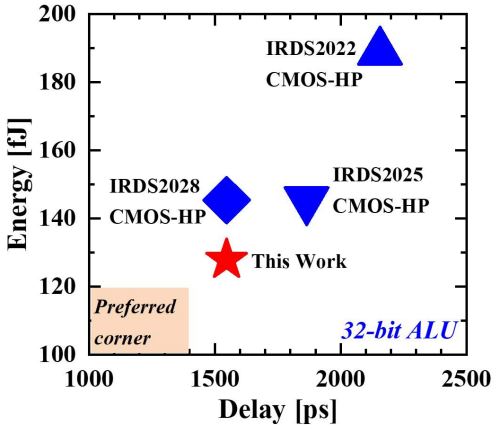

Fig. 8. Switching energy versus switching delay for 32-bit ALU composed of optimized NSFETs in HP condition, compared with 2022, 2025, and 2028 IRDS target.

IV. CONCLUSION

In this work, a machine learning-based MOO workflow for channel geometry and doping profile of three-layer-stacked NSFETs is demonstrated. The adaptive weighted sum theory is implemented in MOO targeting device performance characteristics including SS, on–off ratio, and ON-state current, from which distributions of design parameters, W, H, L_{ov}, and σ are obtained as design windows for NSFETs. An optimized set of parameters is extracted from the distribution and the corresponding NSFETs performance is assessed by TCAD simulation and circuit simulation, the performance of which is compared with IRDS, indicating that the NSFETs with optimal channel geometry and doping profile meet the 2025 and 2028 IRDS targets.

REFERENCES

[1] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *Proc. Symp. VLSI Technol. (VLSIT)*, Jun. 2012, pp. 131–132.
[2] S.-G. Hur *et al.*, "A practical Si nanowire technology with nanowire-on-insulator structure for beyond 10nm logic technologies," in *IEDM Tech. Dig.*, Dec. 2013, p. 26.
[3] N. Loubet *et al.*, "Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET," in *Proc. Symp. VLSI Technol.*, Jun. 2017, pp. T230–T231.
[4] Y. M. Lee *et al.*, "Accurate performance evaluation for the horizontal nanosheet standard-cell design space beyond 7nm technology," in *IEDM Tech. Dig.*, Dec. 2017, p. 29.

[5] D. Yakimets *et al.*, "Power aware FinFET and lateral nanosheet FET targeting for 3nm CMOS technology," in *IEDM Tech. Dig.*, Dec. 2017, p. 20.

[6] C. W. Yeung *et al.*, "Channel geometry impact and narrow sheet effect of stacked nanosheet," in *IEDM Tech. Dig.*, Dec. 2018, pp. 6–28.

[7] J. Zhang *et al.*, "High-k metal gate fundamental learning and multi-V_t options for stacked nanosheet gate-all-around transistor," in *IEDM Tech. Dig.*, Dec. 2017, pp. 1–22.

[8] S. Ramesh *et al.*, "Record performance top-down $In_{0.53}Ga_{0.47}As$ vertical nanowire FETs and vertical nanosheets," in *IEDM Tech. Dig.*, Dec. 2017, pp. 1–17.

[9] Y.-S. Huang *et al.*, "First demonstration of uniform 4-stacked $Ge_{0.9}Sn_{0.1}$ nanosheets with record I_{ON} =73μA at V_{OV}=V_{DS}= -0.5 V and low noise using double $Ge_{0.95}Sn_{0.05}$ caps, dry etch, low channel doping, and high S/D doping," in *IEDM Tech. Dig.*, Dec. 2020, pp. 2–4.

[10] N. Loubet *et al.*, "A novel dry selective etch of SiGe for the enablement of high performance logic stacked gate-all-around NanoSheet devices," in *IEDM Tech. Dig.*, Dec. 2019, pp. 4–11.

[11] A. Agrawal *et al.*, "Gate-all-around strained $Si_{0.4}Ge_{0.6}$ nanosheet PMOS on strain relaxed buffer for high performance low power logic application," in *IEDM Tech. Dig.*, Dec. 2020, p. 2.

[12] S. Mochizuki *et al.*, "Stacked gate-all-around nanosheet pFET with highly compressive strained $Si_{1-x}Ge_x$ channel," in *IEDM Tech. Dig.*, Dec. 2020, pp. 2–3.

[13] J.-P. Colinge *et al.*, "Nanowire transistors without junctions," *Nature Nanotechnol.*, vol. 5, pp. 225–229, Feb. 2010.

[14] M. G. Bardon *et al.*, "Power-performance trade-offs for lateral NanoSheets on ultra-scaled standard cells," in *Proc. IEEE Symp. VLSI Technol.*, Jun. 2018, pp. 143–144.

[15] H. Kim *et al.*, "Optimization of stacked nanoplate FET for 3-nm node," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1537–1541, Apr. 2020.

[16] J. Yao *et al.*, "Physical insights on quantum confinement and carrier mobility in Si, $Si_{0.45}Ge_{0.55}$, Ge gate-all-around NSFET for 5 nm technology node," *J. Electron Devices Soc.*, vol. 6, pp. 841–848, 2018.

[17] Q. Huo *et al.*, "Physics-based device-circuit cooptimization scheme for 7-nm technology node SRAM design and beyond," *IEEE Trans. Electron Devices*, vol. 67, no. 3, pp. 907–914, Mar. 2020.

[18] Q. Huo *et al.*, "A novel general compact model approach for 7-nm technology node circuit optimization from device perspective and beyond," *J. Electron Devices Soc.*, vol. 8, pp. 295–301, 2020.

[19] S.-C. Han and S.-M. Hong, "Deep neural network for generation of the initial electrostatic potential profile," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2019, pp. 1–4.

[20] 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.

[21] S. S. Raju, B. Wang, K. Mehta, M. Xiao, Y. Zhang, and H.-Y. Wong, "Application of noise to avoid overfitting in TCAD augmented machine learning," in *Proc. Int. Conf. Simulation Semiconductor Processes Devices (SISPAD)*, Sep. 2020, pp. 351–354.

[22] K. Mehta and H.-Y. Wong, "Prediction of FinFET current-voltage and capacitance-voltage curves using machine learning with autoencoder," *IEEE Electron Device Lett.*, vol. 42, no. 2, pp. 136–139, Feb. 2021.

[23] C. Akbar, Y. Li, and W. L. Sung, "Machine learning aided device simulation of work function fluctuation for multichannel gate-all-around silicon nanosheet MOSFETs," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5490–5497, Nov. 2021.

[24] H. Carrillo-Nuñez, N. Dimitrova, A. Asenov, and V. Georgiev, "Machine learning approach for predicting the effect of statistical variability in Si junctionless nanowire transistors," *IEEE Electron Device Lett.*, vol. 40, no. 9, pp. 1366–1369, Sep. 2019.

[25] H. Dhillon, K. Mehta, M. Xiao, B. Wang, Y. Zhang, and H. Y. Wong, "TCAD-augmented machine learning with and without domain expertise," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5498–5503, Nov. 2021.

[26] K. Ko, J. K. Lee, M. Kang, J. Jeon, and H. Shin, "Prediction of process variation effect for ultrascaled GAA vertical FET devices using a machine learning approach," *IEEE Trans. Electron Devices*, vol. 66, no. 10, pp. 4474–4477, Oct. 2019.

[27] K. Ko, J. K. Lee, and H. Shin, "Variability-aware machine learning strategy for 3-D NAND flash memories," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1575–1580, Apr. 2020.

[28] J. Wang, Y.-H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, "Artificial neural network-based compact modeling methodology for advanced transistors," *IEEE Trans. Electron Devices*, vol. 68, no. 3, pp. 1318–1325, Mar. 2021.

[29] 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 Devices*, vol. 68, no. 9, pp. 4181–4188, Sep. 2021.

[30] F. Kashfi, S. Hatami, and M. Pedram, "Multi-objective optimization techniques for VLSI circuits," in *Proc. 12th Int. Symp. Quality Electron. Design*, Mar. 2011, pp. 1–8.

[31] T. Wu and J. Guo, "Multiobjective design of 2-D-material-based field-effect transistors with machine learning methods," *IEEE Trans. Electron Devices*, vol. 68, no. 11, pp. 5476–5482, Nov. 2021.

[32] Z. Stanojevic *et al.*, "Physical modeling—A new paradigm in device simulation," in *IEDM Tech. Dig.*, Dec. 2015, pp. 1–5.

[33] C. Fischer *et al.*, "MINIMOS 6 user's guide," Institut Für Mikroelektronik, Technische Universität Wien, Vienna, Austria, 1994.

[34] R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, no. 12, pp. 2192–2193, Dec. 1967.

[35] S. Selberherr, W. Hänsch, M. Seavey, and J. Slotboom, "The evolution of the MINIMOS mobility model," *Solid-State Electron.*, vol. 33, no. 11, pp. 1425–1436, Nov. 1990.

[36] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32. Vancouver, BS, Canada, Dec. 2019, pp. 8026–8037.

[37] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural Networks for Perception*. Amsterdam, The Netherlands: Elsevier, 1992, pp. 65–93.

[38] Z. Zhang, "Improved Adam optimizer for deep neural networks," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–2.

[39] L. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. Autom. Control*, vol. AC-8, no. 1, pp. 59–60, Jan. 1963.

[40] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: New insights," *Struct. Multidisciplinary Optim.*, vol. 41, no. 6, pp. 853–862, Jun. 2010.

[41] I. Y. Kim and O. L. De Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Struct. Multidisciplinary Optim.*, vol. 29, no. 2, pp. 149–158, Feb. 2005.

[42] D. E. Nikonov and I. A. Young, "Benchmarking of beyond-CMOS exploratory devices for logic integrated circuits," *IEEE J. Exploratory Solid-State Comput. Devices Circuits*, vol. 1, no. 1, pp. 3–11, Dec. 2015.