Solid-State Electronics 198 (2022) 108450

Contents lists available at ScienceDirect

Solid State Electronics

journal homepage: www.elsevier.com/locate/sse

A novel methodology for neural compact modeling based on knowledge transfer☆,☆☆

Ye Sle Cha *, Junghwan Park, Chanwoo Park, Soogine Chong, Chul-Heung Kim, Chang-Sub Lee, Intae Jeong, Hyunbo Cho

Research & Development center, Alsemy Inc., 16 Bongeunsa-ro 78-gil, Gangnam-gu, Seoul, 06154, South Korea

ARTICLE INFO

Keywords:
Artificial neural network
Compact modeling
Deep learning
Knowledge transfer
Meta learning
MOSFET
Statistical modeling
Transfer learning

ABSTRACT

This work presents a novel approach of using knowledge transfer to increase the accuracy of artificial neural network (ANN)-based device compact models, or neural compact models. This is useful when the amount of data available for training an ANN is limited. By utilizing relatively abundant data of a previous technology node, physical phenomena that are not evident in the limited data of the target technology node (e.g. gate-induced drain leakage) are accurately predicted. When meta learning algorithms are used, the accuracy of the model significantly increases, with relative linear error 10 times lower compared to the case when prior knowledge is not incorporated. The proposed methodology can be used to model future generation devices with limited data, utilizing data from well-characterized past technology node devices.

1. Introduction

Analytical compact models are currently the most widely used form of device models for circuit simulations. Such models are based on parametric equations derived from device physics and can model the device behavior of the entire operation regime by performing only a few representative measurements. However, only physics manually incorporated in the equations can be modeled with high accuracy. As more complicated physical phenomena are introduced with device scaling, the number of fitting parameters have rapidly increased [1]. This results in parameter fitting becoming an increasingly complicated task, while still not being sufficient to achieve the desired accuracy.

Artificial neural network (ANN)-based device compact models, or neural compact models, have been introduced for faster device model generation with higher accuracy based on measured data. Although it is possible to achieve sufficient accuracy with a large dataset [2], it is costly to obtain such dataset. To overcome this issue, previous work directly incorporated in the ANN model the physics that was already understood [3]. The accuracy improved without increasing the amount of data but the specific physics was explicitly included in the ANN model.

Hence we propose using knowledge transfer methods, which leverage abundant data of a similar device technology to build an accurate neural compact model of the device of interest with a limited amount of data. The relevant physics is automatically incorporated in the model without any hand-crafted modeling effort, achieving high accuracy even with limited data.

2. Knowledge transfer for device modeling

We propose a new modeling framework for tackling the scarcity of data for the device to be modeled, the *target* device. We set a similar environment to parameter extractions for analytical models, where only a few I–V sweeps are measured for a limited number of channel width (W), channel length (L), and temperature (T) combinations (W/L/T) of the target device. Fig. 1 shows the contrast between the available data for ANN training, and the *test data* for evaluating the accuracy of the trained ANN. The amount of the test data is approximately 47 times greater than that of the available data, consisting of 24 I–V sweeps.

The methodology consists of two parts, each part corresponding to the equation development and the parameter extraction, respectively, of the analytical model as in Fig. 2. First, instead of developing an equation-based model, we *pretrain* an ANN to learn the device physics from a *source* device with a large dataset available, equivalent to the union of two datasets in Fig. 1 for each W/L/T. The model parameter

☆ This work (Grants No. S3031427) was supported by Business for Startup growth and technological development (TIPS Program) funded by Korea Ministry of SMEs and Startups in 2022. This work was also supported by WISET-2022-487 in Korea.
☆☆ The review of this paper was arranged by Cristina Medina-Bailon.
* Corresponding author.
E-mail address: yesle.cha@alsemy.com (Y.S. Cha).

https://doi.org/10.1016/j.sse.2022.108450

Available online 14 September 2022

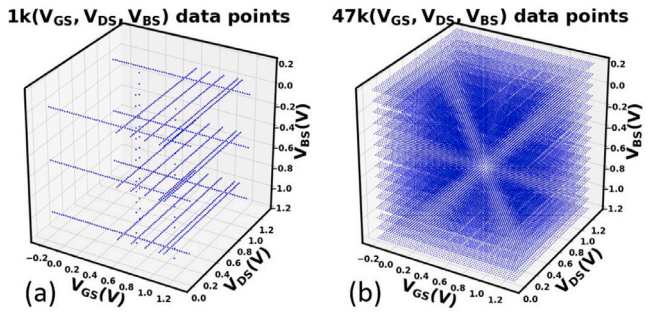


Fig. 1. Bias domain for each W/L/T of the target device of (a) available data for training and (b) test data.

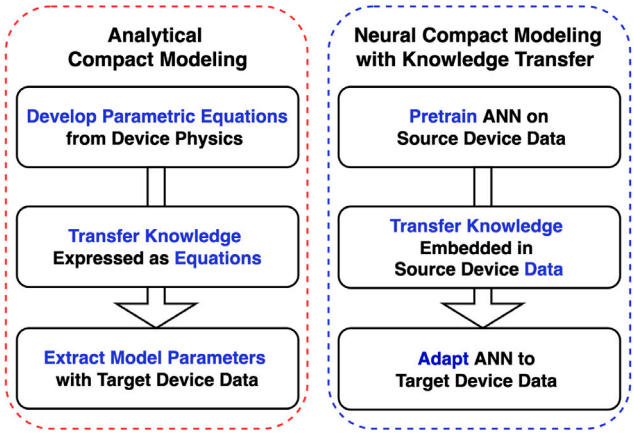


Fig. 2. Comparison of analytical compact modeling and neural compact modeling using knowledge transfer methods.

extraction is also replaced with the *adaptation* of the pretrained ANN to the limited available data of the target device. While the analytical model transfers knowledge expressed in equations, the proposed ANN model transfers knowledge embedded in the source device data. The whole procedure is called "knowledge transfer". Two knowledge transfer methods are discussed in this work — transfer learning and meta learning.

2.1. Transfer learning

Transfer learning aims to leverage knowledge of a related task to enhance the performance of an ANN on a target task [4]. As described in Fig. 3, we select a W/L/T dataset of the source device, and *pretrain* an ANN, a multilayer perceptron (MLP), on that dataset. Assuming that the I–V characteristics of the source and the target devices share similar features [4], we adapt the pretrained ANN to a limited W/L/T dataset of the target device by *fine-tuning* it. All parameters of the ANN are updated by back-propagation to minimize the error between the data and the ANN predictions. The training time for fine-tuning is much shorter than that of pretraining.

2.2. Meta learning

Meta learning focuses on training an ANN to *learn to learn* over multiple learning episodes, so that the trained or *meta-trained* ANN quickly adapts to *unseen* limited data with high test-time accuracy [5]. Fig. 4 illustrates the overall procedures. We employ an encoder–decoder architecture, where the encoder extracts compressed *representations* of W/L/T datasets, and the decoder makes predictions utilizing those representations. We also use a few more MLPs including the updater,

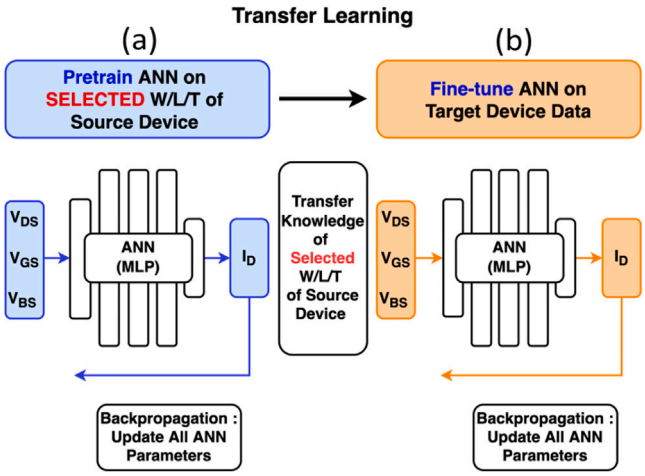


Fig. 3. Illustration of two processes of transfer learning: (a) *pretraining* and (b) *fine-tuning*.

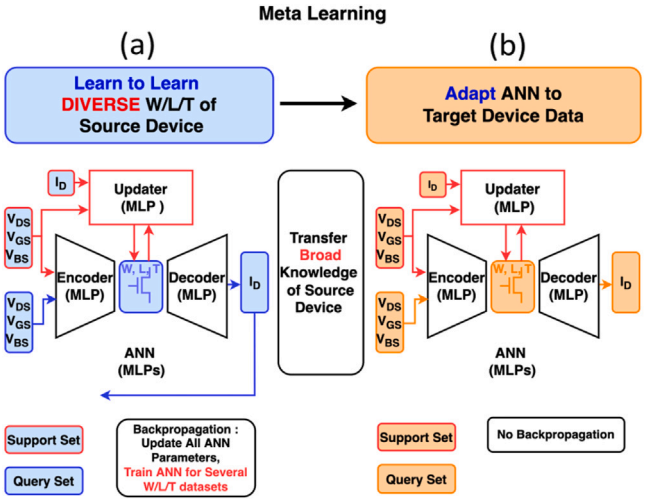


Fig. 4. Detailed description of meta learning processes: (a) *meta-training* and (b) *adaptation*.

to assist the encoder in producing valid representations, by applying MetaFun techniques [6].

During meta-training, several W/L/T datasets of the source device are used, and each of them is parted into disjoint "support set" and "query set" [5]. The relation between the bias voltages and the current in the support set is utilized by the updater to improve the representation (transistor symbol), and that helps the decoder make valid predictions on the query set. The red and blue arrows indicate each process, respectively, in Fig. 4(a). By repeating the above processes on diverse W/L/T datasets, the ANN learns how to quickly adapt to each support set, so that it produces accurate predictions on unseen query set.

Consequently, for a target device similar to the source device, the meta-trained ANN quickly adapts to a limited available dataset, or support set, for *any* W/L/T of the target device, by leveraging the broad knowledge of the source device. On a large test dataset, or query set, the ANN instantiates accurate predictions without fine-tuning, as depicted in Fig. 4(b).

3. Experiments

To test the proposed framework, data generated from SPICE simulation of 45 nm and 32 nm technology node nMOSFET compact

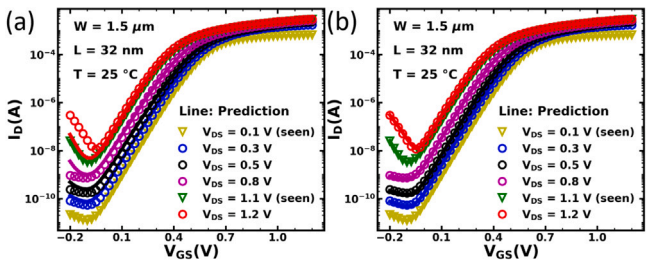


Fig. 5. $I_D - V_{GS}$ ANN predictions (lines), training data (triangles), and test data (circles) of (a) the randomly initialized ANN and (b) the fine-tuned ANN.

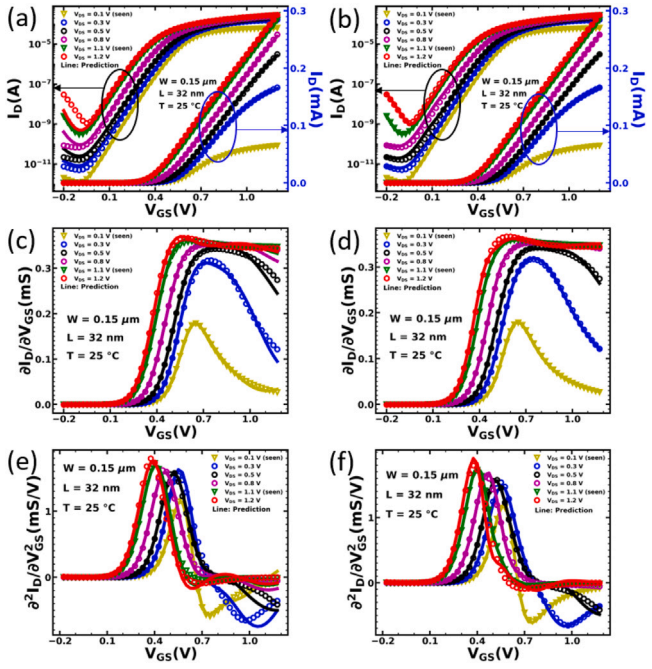


Fig. 6. $I_D - V_{GS}$ and first, second derivative of $I_D - V_{GS}$ ANN predictions (lines), training data (triangles), and test data (circles) of the randomly initialized ANN [(a), (c) and (e), respectively] and of the meta-trained ANN [(b), (d) and (f) respectively].

models [7] are used as source and target device data, respectively. We validate our methodology by comparing it to the case where an ANN is randomly initialized and is trained only on the limited available data of the target device.

First, we apply transfer learning to an ANN and compare its test result with that of a randomly initialized ANN. On our target available data (see Fig. 1(a)), only 6 $I_D - V_{GS}$ sweeps are provided. In Fig. 5, we see that such limited data for training a randomly initialized ANN are not enough to accurately model gate-induced drain leakage (GIDL) currents. In contrast, the ANN with transfer learning successfully predicts GIDL currents for all V_{DS} and V_{BS} considered, by using learned physical knowledge from pretraining stage.

Next, we apply meta learning and compare the predicted I–V characteristics and their higher order derivatives given by the meta-trained ANN and the randomly initialized ANN in Fig. 6. The meta-trained ANN smoothly and accurately predicts the I–V characteristics up to two differentiations without any training on their derivatives as in [8], by effectively utilizing the learned curve characteristics.

Table 1 compares computational costs and test errors for all three ANN training methods. During pretraining, one W/L/T dataset of the source device is used for transfer learning, and 240 such datasets are used for meta learning with longer pretraining time. For adaptation to the target device data, 54 W/L/T datasets are used for all methods.

Table 1
Knowledge Transferred ANN vs. Randomly Initialized ANN.

	Random initialization	Transfer learning	Meta learning
Pretraining time	N/A	646 s	17 h
Adaptation time (per W/L/T)	538 s	186 s	1 s
Relative linear error (%)	22.9	4.3	2.3
Relative log error (%)	1.56	0.40	0.11

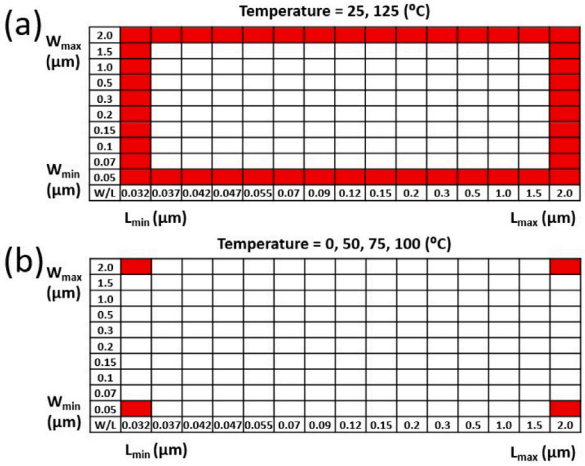


Fig. 7. Domain of W/L/T of 108 target datasets used for testing ANN models for matching the electrical parameters. (a) Domain of W, L for temperature T = 25, 125 ($°C$). (b) Domain of W, L for temperature T = 0, 50, 75, 100 ($°C$).

The meta-trained ANN not only requires the shortest adaptation time but also shows the lowest average relative linear and log errors, which are almost 10 times lower than in the results of the randomly initialized ANN.

Finally, we evaluate the prediction accuracy of electrical parameters, such as channel current at low and high drain biases (I_{DLIN} and I_{DSAT}, respectively), threshold voltage (V_{TH}), and GIDL current. The available data for each W/L/T target dataset are reduced by nearly 36 percent compared to Fig. 1(a), since data near the electrical parameters are excluded. A total of 108 W/L/T datasets, including minimum and maximum values of W/L/T of the target device, is used for testing, as denoted in Fig. 7.

Fig. 8 shows the I_{DLIN} and I_{DSAT} relative linear errors, V_{TH} differences, and GIDL current log differences between the test data and the predicted data using random initialization, transfer learning, and meta learning. The result confirms that the meta-trained ANN captures the most important features of the I–V characteristic for *any* W/L/T of the target device in a much more stable and accurate way than the randomly initialized ANN. The fine-tuned ANN shows good performances, but the variance of the observed errors tends to be higher than that of the meta-trained ANN, since its performance varies depending on the similarity between the source and the target datasets.

4. Conclusion

We have developed a novel framework for neural compact modeling by applying advanced knowledge transfer methods. The resulting model learns the device physics underlying widely available device data and uses that knowledge to predict physically consistent I–V characteristics for any W/L/T of a target device with excellent accuracy, even if the available data of that device are limited.

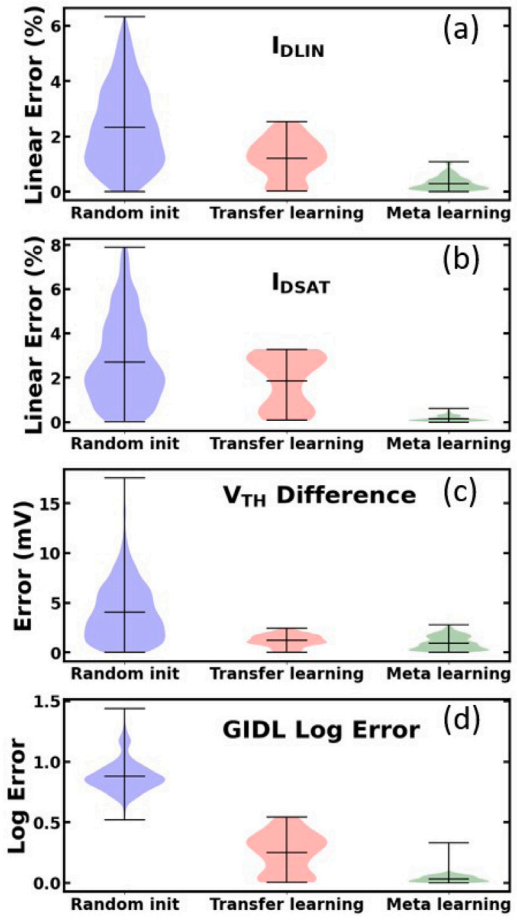


Fig. 8. Comparison of relative linear errors for fitting (a) I_{DLIN} and (b) I_{DSAT}, (c) V_{TH} differences, and (d) log errors for fitting GIDL current by ANN predictions for three methods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] BSIM Research Group, UCBerkeley. BSIM4. 2020, http://bsim.berkeley.edu/models/bsim4/.

[2] Habal H, Tsonev D, Schweikardt M. Compact models for initial MOSFET sizing based on higher-order artificial neural networks. In: 2020 ACM/IEEE 2nd workshop on machine learning for CAD. 2020, p. 111–6.

[3] Kim Y, Myung S, Ryu J, Jeong C, Kim D. Physics-augmented neural compact model for emerging device technologies. In: 2020 international conference on simulation of semiconductor processes and devices. 2020, p. 257–60.

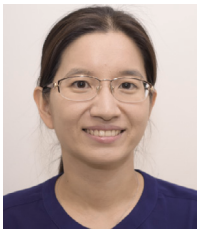
[4] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016, http://www.deeplearningbook.org.

[5] Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-learning in neural networks: A survey. IEEE Trans Pattern Anal Mach Intell 2022;44(9):5149–69.

[6] Xu J, Ton J-F, Kim H, Kosiorek A, Teh Y. MetaFun: Meta-learning with iterative functional updates. In: III HD, Singh A, editors. Proceedings of the 37th international conference on machine learning. Proceedings of Machine Learning Research, 119, PMLR; 2020, p. 10617–27.

[7] Nanoscale Integration and Modeling Group, ASU. PTM high performance 45nm, 32nm MOSFET. 2008, http://ptm.asu.edu/.

[8] 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(3):1318–25.



Ye Sle Cha received her Ph.D. in Mathematics from Stony Brook University, U.S.A. in 2013. She had completed her B.S. in Mathematics at KAIST, South Korea in 2007. She currently works as a Research Scientist in Alsemy Inc., South Korea. Her research interests include scientific machine learning, meta learning, and neural compact modeling.

Junghwan Park received the B.S. degree in Space Science from Kyunghee University, South Korea in 2021. He currently works as a Research Engineer in Alsemy Inc., South Korea.

Chanwoo Park received his B.S. and M.S. in Electrical Engineering from Seoul National University in 2009 and 2012, respectively. He joined Alsemy Inc. in 2020, as a Chief AI Officer. His current research interests include neural compact modeling, physics-based neural network, meta learning, and Bayesian inference.

Soogine Chong received the B.S. degree in Electrical Engineering from Seoul National University in 2003, and the M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA, U.S.A. in 2006 and 2012, respectively. She is currently Chief Scientist with Alsemy Inc., South Korea.

Chul-Heung Kim received the B.S. and Ph.D. degrees in Electrical Engineering from Seoul National University, Seoul, South Korea, in 2013 and 2019, respectively. He is currently Planning Manager with Alsemy Inc., South Korea.

Chang-Sub Lee received his B.S degree in Electrical and Electronic Engineering from KAIST in 1991, and the M.S. degree in Semiconductor Engineering from Sungkyunkwan University, South Korea in 2002. He is currently Research Fellow with Alsemy Inc., South Korea.

Intae Jeong received the B.S. degree in Mechanical Engineering, the M.S. degree in Physics and the Ph.D. degree in Nano Science from Seoul National University, South Korea in 2007, 2009 and 2014, respectively. He is currently Chief Product Officer with Alsemy Inc., South Korea.

Hyunbo Cho is founder and CEO of Alsemy Inc. Previously, he received the B.S. degree in Electrical Engineering from Seoul National University in 2005, and the M.S. degree in Electrical Engineering from Stanford University, Stanford, CA, U.S.A. in 2007.