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 and Hyunbo Cho
Research & Development Center, Alsemy Inc., Seoul, Korea. Email: yesle.cha@alsemy.com

***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.**

***Index Terms*—Artificial neural network, compact modeling, deep learning, knowledge transfer, meta learning, MOSFET, statistical modeling, transfer learning.**

I. Introduction

Artificial neural network (ANN)-based device compact models, or neural compact models, use machine learning to model devices for circuit simulation based on data [2]. However, it is often difficult to obtain a large dataset from novel devices, which limits the accuracy of the neural compact model. To overcome this issue, previous work directly incorporated in the ANN model the physics that was already understood [1]. In this paper, we propose knowledge transfer methods in which the relevant physics in the data is automatically included in the model, without the need to understand such physics.

II. Knowledge Transfer For Device Modeling

The objective of this paper is to provide a new modeling framework for tackling the scarcity of data for a target device. In this framework, we set a similar environment to the case when fitting parameters are extracted for analytical models (e.g. BSIM), where only few I-V sweeps are measured for each channel width (W), channel length (L), and temperature (T) of the target device. Fig. 1 shows the contrast between available data, which are used for ANN training, and test data, which are used to examine the accuracy of the final ANN.

The methodology consists of two parts. First, we *pretrain* the ANN to learn device physics from largely available planar MOSFET data. For each W, L, T, a large amount of data, on par with the union of two data shown in Fig. 1, is used to capture the delicate physical phenomena. Second, we *adapt* the pretrained ANN to the limited target device data, so that the learned physics becomes consistent with the target device. The whole procedure is called 'knowledge transfer'. The details of how each part is conducted depend on the knowledge transfer method used, as explained in Fig. 2.

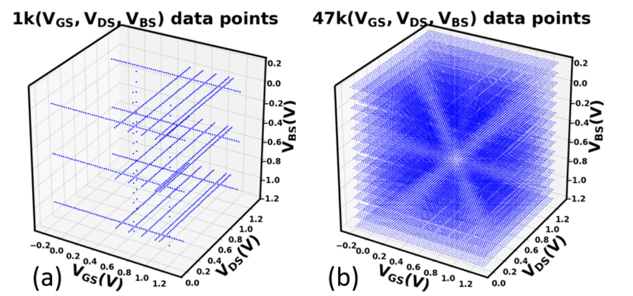

Fig. 1. Bias domain for each W, L, T of the target device of (a) available data, consisting of 24 current-voltage sweeps, and (b) test data with approximately 47 times more data points than those of the available data.

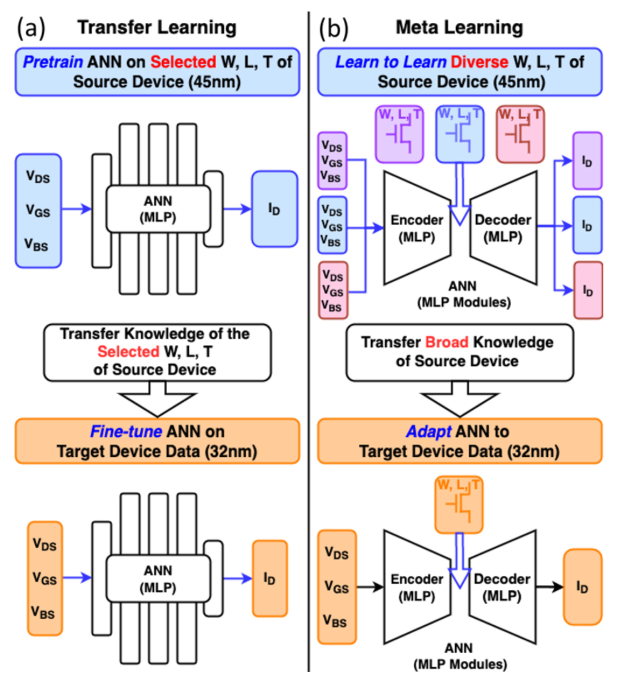

Fig. 2. (a) For transfer learning, we *pretrain* a multilayer perceptron (MLP) on a selected W, L, T dataset of the source device, then *fine-tune* that ANN on limited W, L, T data of the target device. (b) For meta learning, the ANN *learns to learn* by being *meta trained* to repeatedly make accurate predictions utilizing the condensed information on each W, L, T data of the source device that the encoder extracts (represented by transistor symbols). The ANN is adapted to the target device data by extracting relevant information with limited W, L, T data (transistor symbol), and instantiates predictions (in black).

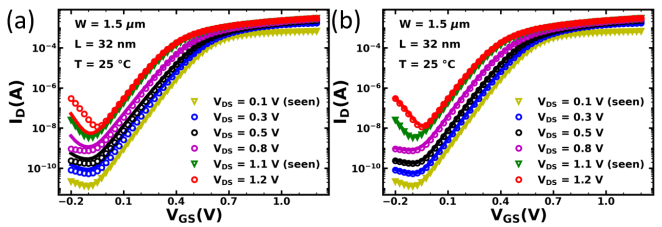

Fig. 3. I_D-V_{GS} ANN predictions (lines), training data (triangles), and test data (circles) of (a) the randomly initialized ANN and (b) the fine-tuned ANN, showing the fine-tuned ANN accurately modeling GIDL.

TABLE I
KNOWLEDGE TRANSFERED ANN VS. RANDOMLY INITIALIZED ANN

	Random Initialization	Transfer Learning	Meta Learning
Pretraining Time	N/A	646 sec.	17 hours
Adaptation Time (per W, L, T)	538 sec.	186 sec.	1 sec.
Relative Linear Error (%)	22.9	4.3	2.3
Relative Log Error (%)	1.56	0.40	0.11

While transfer learning transfers knowledge of a selected W, L, T dataset of the source device, with meta learning, we aim to build a machinery which effectively passes broad knowledge of *all* W, L, T datasets of the source device. In meta learning, the ANN *learns to learn* by performing several learning episodes with diverse W, L, T datasets as described in Fig. 2. After meta training is completed, the ANN quickly leverages the information from limited data of *any* W, L, T of the target device to instantiate accurate predictions on large test data. We also apply MetaFun techniques [3] to capture subtle interactions between data points.

III. EXPERIMENTS

To test the proposed framework, 45 nm and 32 nm technology node MOSFETs [4] are used as source and target devices, respectively. Data for experiments are generated by SPICE simulations. We validate our methodology by comparing it to the case where the ANN is randomly initialized and trained solely on the limited data of the target device.

In Fig. 3, in contrast to the randomly initialized ANN, the fine-tuned ANN successfully predicts gate-induced drain leakage (GIDL) current by using previously learned physical knowledge. Similarly, in Fig. 4, the meta trained ANN predicts I-V characteristics more accurately for up to two differentiations compared to the randomly initialized ANN, without any pre-imposed conditions on derivatives.

Table I compares computational costs and test errors for all three ANN training methods. During pretraining, one W, L, T dataset is used for transfer learning, and 240 such datasets are used for meta learning, with a long training time. For adaptation to the target device data, 52 W, L, T datasets are used for all methods. The meta trained ANN shows the lowest average test errors with the shortest adaptation time.

In Fig. 5, the test results for matching electric parameters such as I_{DLIN}, I_{DSAT}, V_{TH}, and GIDL current are shown for each ANN training method. The meta trained ANN captures the electric parameters for *any* W, L, T in a much more stable

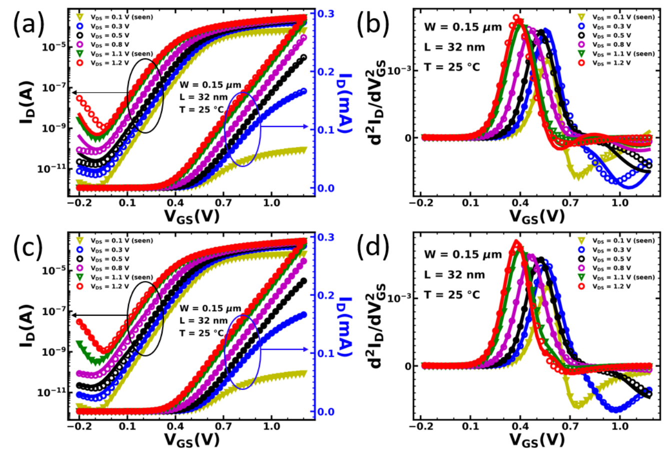

Fig. 4. I_D-V_{GS} and second derivative of I_D-V_{GS} ANN predictions (lines), training data (triangles), and test data (circles) of the randomly initialized ANN [(a) and (b), respectively], and of the meta trained ANN [(c) and (d), respectively], showing the meta trained ANN achieving significantly better fitting compared to the randomly initialized ANN.

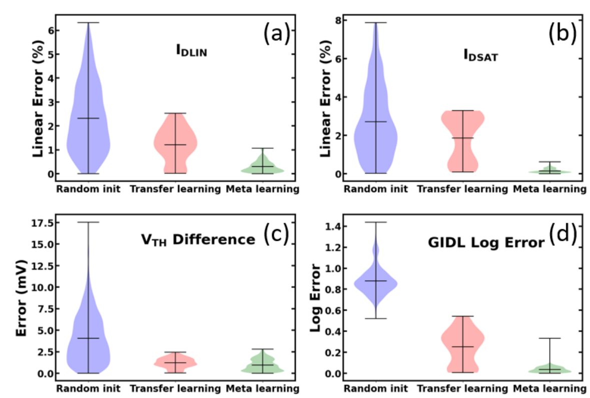

Fig. 5. Comparison of relative linear errors for fitting (a) I_{DLIN} and (b) I_{DSAT}, (c) V_{TH} differences, and (d) log errors for fitting GIDL current by ANN predictions for three methods, showing improved model accuracy for knowledge transfer methods compared to random initialization. For target device data, we select 108 well-balanced W, L, T datasets and reduce the available data for each W, L, T by nearly 36 percent compared to Fig. 1 (a).

and accurate way compared to the randomly initialized ANN. The fine-tuned ANN shows good performance, but its limit lies on higher variance.

IV. CONCLUSION

We develop a novel framework for neural compact modeling by applying advanced knowledge transfer techniques. The resulting model learns the device physics underlying widely available planar MOSFET 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.

REFERENCES

[1] Y. Kim et al., *SISPAD*, pp. 257-260, 2020.
[2] J. Wang et al., *IEEE Trans Electron Devices*, vol. 68, no. 3, pp. 1318-1325, 2021.
[3] J. Xu et al., In *Proceedings of the 37th International Conference on Machine Learning*, pp. 10617-10627, 2020.
[4] Predictive Technology Model (PTM), http://ptm.asu.edu/