FuncAnoDe: A Function Level Anomaly Detection in Device Simulation

Tae IL Oh[1*], Hong Chul Nam[1,2*], Chanwoo Park[1], Hyunbo Cho[1]
[1]Research & Development Center, Alsemy Inc., Seoul, Korea
[2]Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland

2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) | 979-8-3315-1635-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/SISPAD62626.2024.10733289

***Abstract*—In semiconductor device simulations, the reliance on empirical compact models, such as the Berkeley Short-channel IGFET Model (BSIM) and neural compact models, introduces approximations that may significantly diverge from actual physical phenomena. Identifying and filtering out unphysical behaviors and erroneous simulation outcomes is a challenging task, traditionally requiring extensive expert involvement and incurring high costs. In response, we introduce FuncAnoDe, a novel neural operator for unsupervised functional anomaly detection in semiconductor simulation datasets. FuncAnoDe is the first to offer deep learning-based function-level anomaly detection without manual expert intervention. Its function-level encoder-decoder architecture enables applications across a diverse range of device parameters and simulations, ensuring scalability and high accuracy in identifying physically implausible parameter configurations. Our evaluations were conducted through complex capacitance-voltage (C-V) curve analysis, and FuncAnoDe demonstrated its effectiveness in anomaly detection by achieving a 100.00% accuracy without reliance on manual labeling. FuncAnoDe provides a methodological advancement that enhances the precision, reliability, and efficiency of semiconductor design and simulation workflows.**

***Index Terms*—Anomaly Detection, Deep Learning, Neural Operator, Device Modeling**

I. Introduction

In semiconductor device simulation, ensuring the validity of models is crucial for advancing technology. The use of empirical compact models, such as the Berkeley Short-channel IGFET Model (BSIM), and emerging neural compact models [18], has become standard practice. These models have significantly advanced semiconductor technology but rely on approximations of physical phenomena. Consequently, discrepancies can arise between the behaviors predicted by simulations and those observed in real devices.

On the other hand, the application of large-scale training techniques to device simulations is increasing. In [3], pre-trained models utilizing meta-learning techniques are introduced, while [12] explores the use of Generative Artificial Intelligence (GenAI) for Monte Carlo simulations. Furthermore, [11] introduces the first large-scale pre-trained foundation models for I-V and C-V device modeling.

These advancements highlight the growing importance of rigorous data validation processes. However, the current dependence on expert manual inspections of training data is not scalable and demands substantial resources, posing challenges to ensure the robustness and reliability of simulation models.

In response to these challenges, we introduce FuncAnoDe, a novel approach designed to automate and streamline the process of anomaly detection in semiconductor device simulations. FuncAnoDe leverages a function autoencoder (AE) [7] using an attention operator [19] and perceiver-like architecture [5]. This approach enables the efficient identification of anomalies through reconstruction loss metrics without requiring manual labeling. To our knowledge, FuncAnoDe is the first neural operator to apply unsupervised deep learning for anomaly detection in semiconductor device simulations. Based on our extensive experiments, it demonstrates 100% accuracy in detecting anomalies across simulations of 32nm technology node devices, while only pre-trained on larger-scale device data without labels. This advancement demonstrates FuncAnoDe's potential to significantly improve the accuracy and reliability of device models.

*Equal contribution. Email: {taeil.oh, hongchul.nam}@alsemy.com
Corresponding author: Hyunbo Cho. Email: hyunbo.cho@alsemy.com

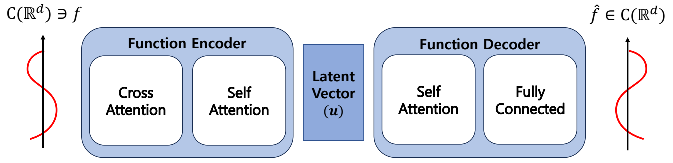


Fig. 1: **Model Architecture:** Function-encoder encodes function f into a latent vector u and Function-decoder decodes the latent vector into the reconstructed function $\hat{f}$

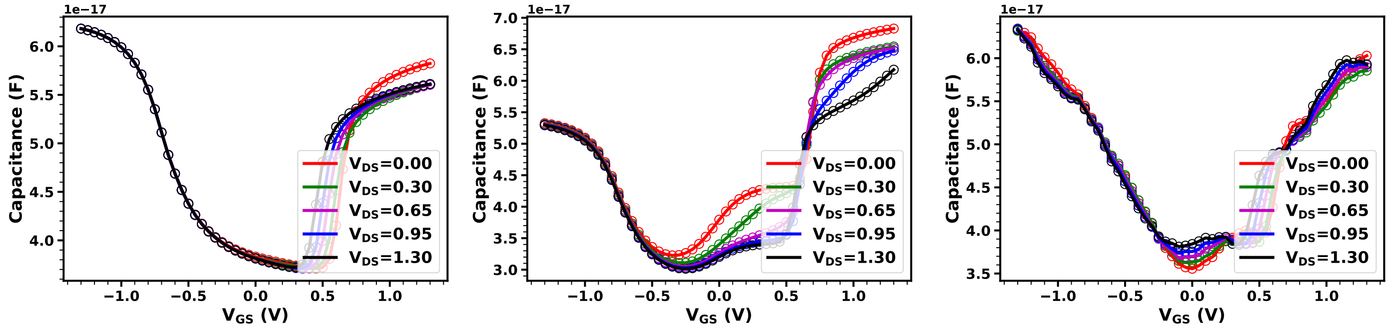
In summary, FuncAnoDe makes the following contributions:

- FuncAnoDe is the first neural operator to apply unsupervised deep learning techniques for anomaly detection in semiconductor device simulations.
- FuncAnoDe extends the AE to handle infinite-dimensional functional data such as C-V curves for anomaly detection in device simulations.
- FuncAnoDE achieves near-perfect accuracy in detecting anomalies in unseen target technology devices.

II. FuncAnoDe: Functional Anomaly Detection

FuncAnoDe is an autoencoder architecture based on attention neural operators [1]. Neural operators [6] are specialized neural networks that learn mappings between two infinite-dimensional function spaces. They possess a discretization invariance property, guaranteeing the convergence to the true solution while allowing evaluations over arbitrary geometries and discretization sizes. This capability is particularly advantageous when dealing with measurement data of varying resolutions. Additionally, neural operators can generalize to irregular, non-uniform meshes, making them practical for real-world measurement data in arbitrary sequences.

Attention neural operators [1] rigorously extend attention [19] to function spaces. In addition to the properties of neural operators, attention neural operators enable the parallel processing of data sequences over irregular grids and improve the representation of complex interactions. The attention operator takes a query $q \in \mathbb{R}^{d_q}$, key $k \in \mathbb{R}^{d_k}$, and value $v \in \mathbb{R}^{d_v}$, and it can be divided to two types, namely cross-attention and self-attention. Cross attention (CA) restricts the query q to be one vector and key and value to be another, i.e. $k = v$. It establishes a relationship between two data sequences. Given a



(a) **Regular Simulation** (b) **Simulated Anomaly** (c) **Augmented Anomaly**

Fig. 2: **Examples of Each Data Type.** (a) Original test dataset generated by human experts with BSIM. (b) Anomaly introduced by human error with incorrect parameter input to BSIM. (c) Anomaly created by removing random frequencies from the original dataset (a).

set of learnt latent vectors $\{u_i\}_{i=1}^{N_u}$ with $u_i \in \mathbb{R}^{d_u}$ and an input function sequence $\{(x_j, f(x_j))\}_{j=1}^{N_x}$, we define

$$CA(q = \{u_i\}_{i=1}^{N_u}, k = v = \{(x_j, f(x_j))\}_{j=1}^{N_x}) = \{u_i'\}_{i=1}^{N_u}, \quad (1)$$

where u_i' represents learned latent vectors incorporating function sequence information. On the other hand, self-attention (SA) uses the same data sequence for queries, keys, and values, i.e. $q = k = v$, and computes the interaction within the same data sequence. Under the same assumptions as cross-attention, we define

$$SA(q = k = v = \{u_i\}_{i=1}^{N_u}) = \{u_i'\}_{i=1}^{N_u}, \quad (2)$$

where u_i' are improved latent vectors.

Motivated by [7], FuncAnoDe utilizes attention neural operators. It initially employs cross-attention to learn the relations between learnable latent vectors and input function evaluations, using C-V simulation data of arbitrary data length. Subsequently, multiple layers of self-attentions are applied over the learned query points to further improve and compress the information of function in these latent vectors. Finally, we apply fully connected layers by concatenated query points for evaluations with the latent vectors for reconstruction.

We train FuncAnoDe on normal datasets from simulations such that

$$\phi^*, \psi^* = \arg\min_{\phi,\psi} \mathbb{E}[\mathcal{D}_\psi \circ \mathcal{E}_\phi \circ f - f], \quad (3)$$

where f are device models randomly sampled from the dataset and $\mathcal{E}, \mathcal{D}$ are encoders and decoders respectively. Refer to Figure 1 for the architecture of the function AE. After training, the network is applied to a dataset consisting of both normal and abnormal datasets, yielding reconstruction errors for each data. Reconstruction errors are then sorted from smallest to largest. Using Kernel Density Estimation (KDE) to estimate the probability density function of the reconstruction errors, along with Silverman's rule of thumb [15], we automatically determine an optimal threshold that distinguishes between low and high reconstruction errors. Datasets with reconstruction errors surpassing the threshold are classified as abnormal, whereas those with errors below the threshold are categorized as normal. We expect that the test abnormal dataset lies beyond the learned data manifold, thus yielding high reconstruction errors.

III. Related Works

Anomaly detection using the reconstruction loss of AE is a widely used technique. For instance, [24, 25] utilize augmentation masks reconstructed using AE. [9] employs Long Short-Term Memory Network to model time series inputs for reconstruction. An energy model, which outputs low scores on trained data regimes, is also explored sx[22]. Generative models like Generative Adversarial Networks or Diffusion Models are also employed for reconstruction-based anomaly detection [13, 20]. However, these methods primarily focus on fixed grid images, time series, or tabular data, whereas FuncAnoDe extends to infinite-dimensional function spaces and is specifically designed to detect anomalies in device data.

TABLE I: Hyperparameters used for training the networks.

Hyperparameters	Values
Number of Latents (N_u)	256
Latent Dimension (d_u)	32
Learning Rate	1e-4
Batch Size	64
Epochs	3000
Optimizer	Adam

IV. Experiments

We compared FuncAnoDe with conventional methods, including Functional Principal Component Analysis (FPCA) [14], as well as clustering techniques like K-means and Spectral Clustering [10, 21]. Since clustering methods struggle to transfer knowledge among different technology nodes, we directly applied clustering to 32nm test datasets. In contrast, FPCA and FuncAnoDe were trained across a broader spectrum of technology scales, eliminating the necessity for training on test data.

A. Training Details

Table I describes the hyperparameters used for training our model. We explain in the following two crucial hyperparameters introduced in Perceiver architecture [5, 7] that define the size of the latent vector u:

- **Number of Latents (N_u):** This determines the number of latent vectors to represent the function.
- **Latent Dimension (d_u):** This specifies the dimension of each latent vector.

We discover that these two hyperparameters are critical for the model's performance because they represent how much function information is compressed in the latent space. They were carefully selected using a grid search to ensure that the architecture accurately captures the intricate features of our device dataset while balancing computational complexity.

B. Datasets

We generated training and test datasets by SPICE simulations [2, 16], utilizing previous technology nodes ranging from 55nm to 180nm for training, and 32nm for testing. This approach aligns with common practices, where simulation and

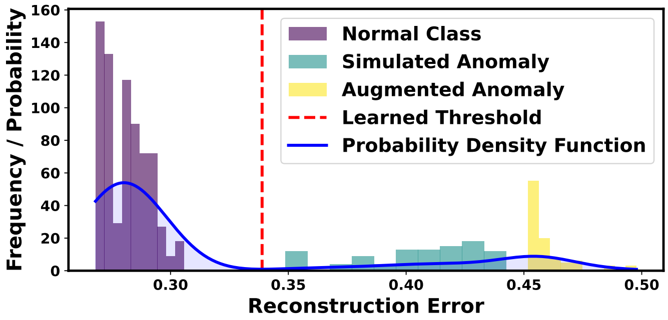


Fig. 3: **Probability density function** fitted to the histogram of reconstruction errors. The threshold (red line) is automatically determined by the fitted density (blue line) using KDE.

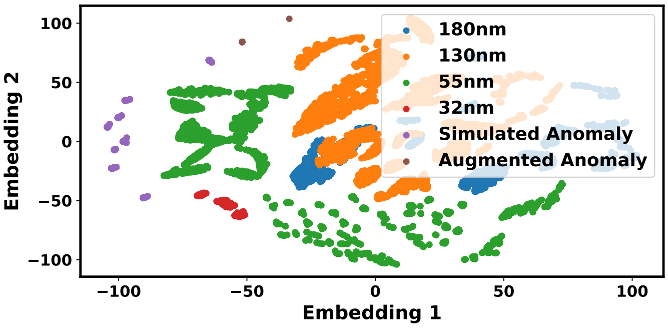


Fig. 4: **t-SNE plot** of latent vectors from the dataset. High-dimensional vectors are projected onto a 2D plot, preserving their relative distances in the original high-dimensional space.

measurement datasets of previously developed technology node devices are abundant.

For anomaly detection, our test dataset comprises three distinct types. Firstly, we have regular simulations (refer to Figure 2a), representing normal C-V curves. Secondly, we incorporate simulated anomalies (refer to Figure 2b), where our device experts manually induced failures in the BSIM model. Lastly, we introduce augmented anomalies created by masking high-frequency modes (refer to Figure 2c), aimed at diversifying anomaly types. The diversity of test data allows us to evaluate our anomaly detection method in various scenarios thoroughly.

V. Results

In this section, we compare the results of FuncAnoDe with those of our baseline models. Table II shows a performance comparison among FPCA, K-means, spectral clustering, and FuncAnoDe. Notably, FPCA yields suboptimal results in anomaly detection. Unlike deep neural networks, which excel in capturing intricate features, FPCA encounters difficulties in anomaly detection due to significant variations in functions sharing similar underlying physics. This variability hinders FPCA's ability to generalize across different types of function data, thereby restricting its effectiveness in identifying anomalies. Furthermore, it requires a fixed grid as input, further limiting its scope of usage.

Across all metrics (accuracy, recall, and precision), FuncAnoDe achieves superior results compared to our baseline methods. Notably, FuncAnoDe achieves 100% recall, indicating its capability to detect anomalies effectively without false negatives. Moreover, it outperforms baseline models by almost double, highlighting its robustness.

To further assess its robustness, we introduced noise to the test data. At 20dB noise level, FuncAnoDe's performance only experiences a moderate 3% drop on average metrics,

TABLE II: Comparison of Performance Metrics

Method	Performance Metrics		
	Accuracy	Recall	Precision
FPCA	10.00%	50.00%	11.25%
K-means Clustering	79.73%	49.50%	49.75%
Spectral Clustering	82.15%	50.00%	56.50%
FuncAnoDe	**100.00%**	**100.00%**	**100.00%**
FuncAnoDe (w. noise 20dB)	98.69%	94.00%	98.75%

demonstrating its resilience in handling realistic data with noise present. We find that under reasonable noise levels, the overall perturbation on the test data has only minor impacts on the reconstruction errors. Refer to Figure 5 for an example of the noisy data.

Figure 3 illustrates the distribution of reconstruction errors, with KDE to fit the probability density function. FuncAnoDe automatically determines the threshold for decision-making, simplifying its use for non-AI experts. Additionally, we observe that FuncAnoDe appears to cluster certain types of anomalies. While this aspect remains as future research, the potential for clustering anomaly types suggests the algorithm could offer enhanced insights for human inspection.

Figure 4 presents a 2D projected plot of latent vectors of function inputs. Intermediate features in the network, as highlighted by [4] and [17], play a critical role in detecting anomalies. Hence, investigating it will open up future development of the method. Despite being trained solely on device curves from 180nm to 55nm, the network effectively clusters anomalies from normal samples based on their latent vectors. This proves that FuncAnoDe can successfully learn meaningful latent features.

VI. Limitations and Future Work

There are a few limitations of the work. First, reconstruction errors could be regarded as 1D projections from function space. Due to the disparity of dimension size between infinite-dimensional function space and 1D real space, it can only compress a limited amount of function information. Second, we can think of reconstruction loss as an energy-based model [23]. Using the energy-based model framework, we can improve our model to more efficiently separate negative samples from positive ones. Third, while the model can distinguish anomalies from normal datasets, it cannot perform clustering of multiple anomaly types, if anomaly types increase.

As a result, it becomes important to extend FuncAnode in the future to support multiple metrics other than reconstruction errors to have a high-dimensional projection measuring the similarity between the original function and reconstructed functions such that one can perform clustering methods for distinguishing diverse types of anomalies. Moreover, we could also extend our work to an energy-based operator by using function-level Langevin dynamics [8] to sample negative functions considered anomalies to further distinct positive and negative features.

VII. Conclusion

This paper presents FuncAnoDe, a function level anomaly detection to detect abnormal simulation results with minimal domain knowledge. To the best of our knowledge, this is the first paper to introduce deep learning based function-level anomaly detection in device simulations. Our comprehensive experimental results demonstrate that our FuncAnoDe outperforms baseline existing unsupervised clustering techniques. Despite not being trained on the test dataset, FuncAnoDe shows robust

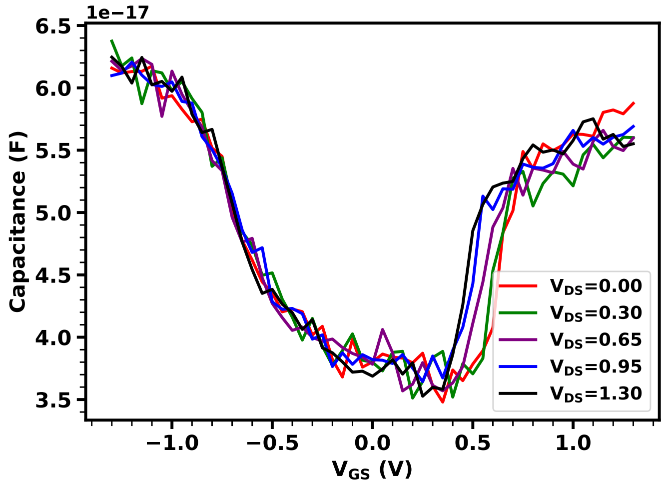


Fig. 5: **Example of noisy input with 20dB noise level** The network successfully detects both normal and abnormal samples even in the presence of such noise, showcasing the robustness of our method in handling potentially noisy real measurements.

generalization capabilities, effectively discriminating physically inaccurate curves.

References

[1] Edoardo Calvello, Nikola B Kovachki, Matthew E Levine, and Andrew M Stuart. Continuum attention for neural operators. *arXiv preprint arXiv:2406.06486*, 2024.

[2] Yu Cao, Asha Balijepalli, Saurabh Sinha, Chi-Chao Wang, Wenping Wang, Wei Zhao, et al. The predictive technology model in the late silicon era and beyond. *Foundations and Trends® in Electronic Design Automation*, 3(4):305–401, 2010.

[3] Ye Sle Cha, Junghwan Park, Chanwoo Park, Soogine Chong, Chul-Heung Kim, Chang-Sub Lee, Intae Jeong, and Hyunbo Cho. A novel methodology for neural compact modeling based on knowledge transfer. *Solid-State Electronics*, 198:108450, 2022.

[4] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022.

[5] Andrew Jaegle, Felix Gimeno, Andy Brock, Oriol Vinyals, Andrew Zisserman, and Joao Carreira. Perceiver: General perception with iterative attention. In *International conference on machine learning*, pages 4651–4664. PMLR, 2021.

[6] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

[7] Seungjun Lee and Taeil Oh. Inducing point operator transformer: A flexible and scalable architecture for solving pdes. In *In Proc. AAAI*, volume 38, pages 153–161, 2024.

[8] Jae Hyun Lim, Nikola B Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, et al. Score-based diffusion models in function space. *arXiv preprint arXiv:2302.07400*, 2023.

[9] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.

[10] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in NeurIPS*. MIT Press, 2001.

[11] Chanwoo Park, Seungjun Lee, Junghwan Park, Kyungjin Rim, Jihun Park, Seonggook Cho, Jongwook Jeon, and Hyunbo Cho. Large-scale training in neural compact models for accurate and adaptable mosfet simulation. *IEEE Journal of the Electron Devices Society*, pages 1–1, 2024.

[12] Chanwoo Park, Hong Chul Nam, Jihun Park, and Jongwook Jeon. Flowsim: An invertible generative network for efficient statistical analysis under process variations. In *Proc. SISPAD*, pages 157–160. IEEE, 2023.

[13] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.

[14] Heng Tao Shen. Principal component analysis. *Encyclopedia of database systems*, 2136, 2009.

[15] Bernard W Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.

[16] James E Stine, Ivan Castellanos, Michael Wood, Jeff Henson, Fred Love, W Rhett Davis, Paul D Franzon, Michael Bucher, Sunil Basavarajaiah, Julie Oh, et al. Freepdk: An open-source variation-aware design kit. In *2007 IEEE international conference on Microelectronic Systems Education (MSE'07)*, pages 173–174. IEEE, 2007.

[17] Hugo Thimonier, Fabrice Popineau, Arpad Rimmel, and Bich-Liên Doan. Beyond individual input for deep anomaly detection on tabular data. *arXiv preprint arXiv:2305.15121*, 2023.

[18] Chien-Ting Tung, Ming-Yen Kao, and Chenming Hu. Neural network-based and modeling with high accuracy and potential model speed. *IEEE Trans. Electron Devices*, 69(11):6476–6479, 2022.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[20] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. Diffusion models for medical anomaly detection. In *International Conference on Medical image computing and computer-assisted intervention*, pages 35–45. Springer, 2022.

[21] Junjie Wu. *Advances in K-means clustering: a data mining thinking*. Springer Science & Business Media, 2012.

[22] Sangwoong Yoon, Young-Uk Jin, Yung-Kyun Noh, and Frank Park. Energy-based models for anomaly detection: A manifold diffusion recovery approach. *Advances in Neural Information Processing Systems*, 36, 2024.

[23] Sangwoong Yoon, Yung-Kyun Noh, and Frank Park. Autoencoding under normalization constraints. In *International Conference on Machine Learning*, pages 12087–12097. PMLR, 2021.

[24] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021.

[25] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021.