IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 December 2023, accepted 2 January 2024, date of publication 9 January 2024,
date of current version 16 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351802

APPLIED RESEARCH

Application of Generative Adversarial Networks for Virtual Silicon Data Generation and Design-Technology Co-Optimization: A Study on WAT and CP

SHIH-NUNG CHEN[1] AND SHI-HAO CHEN[2]
[1]Department of Information Communication, Asia University, Taichung 413305, Taiwan
[2]DigWise Technology Corporation, Hsinchu 302047, Taiwan

Corresponding author: Shih-Nung Chen (nung@asia.edu.tw)

ABSTRACT This study explores the application of Generative Adversarial Networks (GANs) for generating wafer-level Wafer Acceptance Test (WAT) and Chip Probe (CP) test data in chip manufacturing processes, with a focus on Design-Technology Co-Optimization (DTCO). The generated virtual silicon data encompasses essential performance characteristics, physical electrical properties, wafer-level process parameter distributions, as well as implicit information about wafer-level uniformity and defects. This information represents non-random features on wafers, such as similar distorted wafer surfaces observed in lots produced from various batches. This innovative approach overcomes data acquisition barriers, efficiently compresses large datasets while ensuring data confidentiality, and holds immense potential for the development of advanced Electronic Design Automation (EDA) tools, enabling the synergistic optimization of manufacturing processes and chip design flow.

INDEX TERMS Generative adversarial network (GAN), wafer acceptance test (WAT), chip probe (CP), design-technology co-optimization (DTCO), virtual silicon data, electronic design automation (EDA).

I. INTRODUCTION

To attain a competitive advantage in the semiconductor market, chip design processes must aim to enhance production efficiency, reduce costs, and maintain high-quality standards. In recent years, the industry has transitioned from primarily focusing on yield to considering overall productivity. Productivity encompasses a wide range of factors, including cost, price, performance, yield, and competitiveness. Therefore, comprehensive optimization is essential across various aspects, such as design margins, timing sign-off, process recipes, packaging testing, binning strategies, and system-level considerations.

The Design-Technology Co-Optimization (DTCO) methodology has gained significant attention and adoption in physical design processes to enhance the overall productivity and competitiveness of semiconductor chips. In this context, DTCO can be compared to a large-scale neural network optimization process, as shown in Figure 1. The result of our focus on inference is the optimization of productivity, including chip monitoring, Wafer Acceptance Test (WAT), Chip Probe (CP), and System-Level Test (SLT) processes, feature correlation analysis, machine learning, and compensation techniques for binning strategies. Conversely, during the back-propagation optimization phase, our emphasis is on optimizing design and process recipes. This entails chip model calibration, fine-tuning process parameters, timing extraction, customization, and optimization using WAT measurements for the device library. Additionally, it involves On-Chip Variation (OCV) regression for handling local variations and optimizing design margins and sign-off strategies. Figure 2 illustrates the actual implementation details of DTCO proposed in this study.

The associate editor coordinating the review of this manuscript and approving it for publication was Ludovico Minati.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

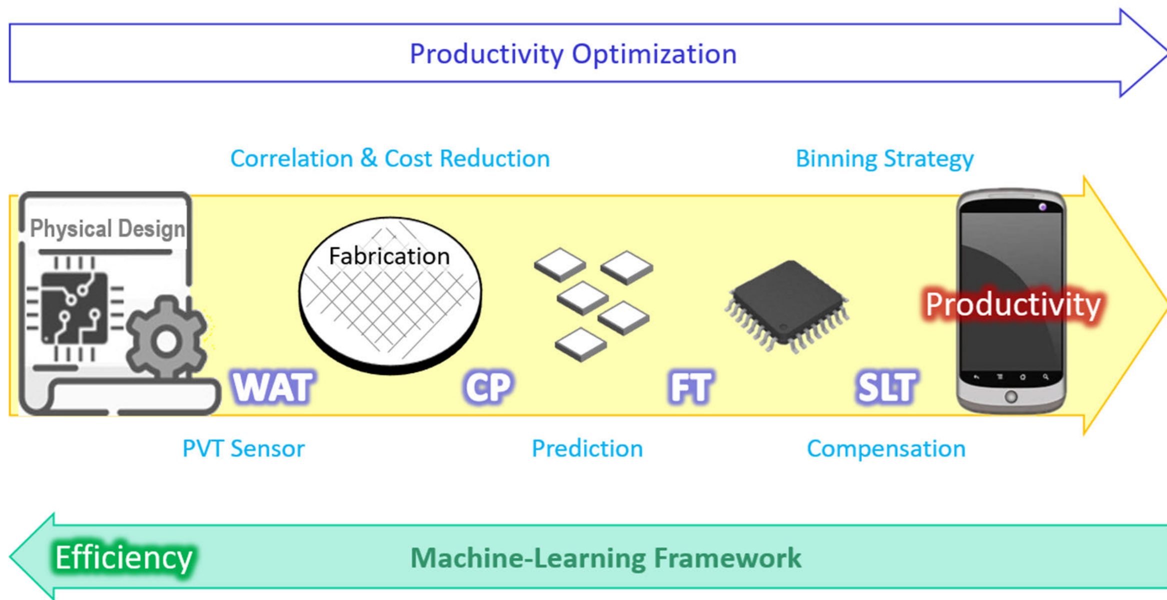

FIGURE 1. Design-Technology Co-Optimization (DTCO) flow.

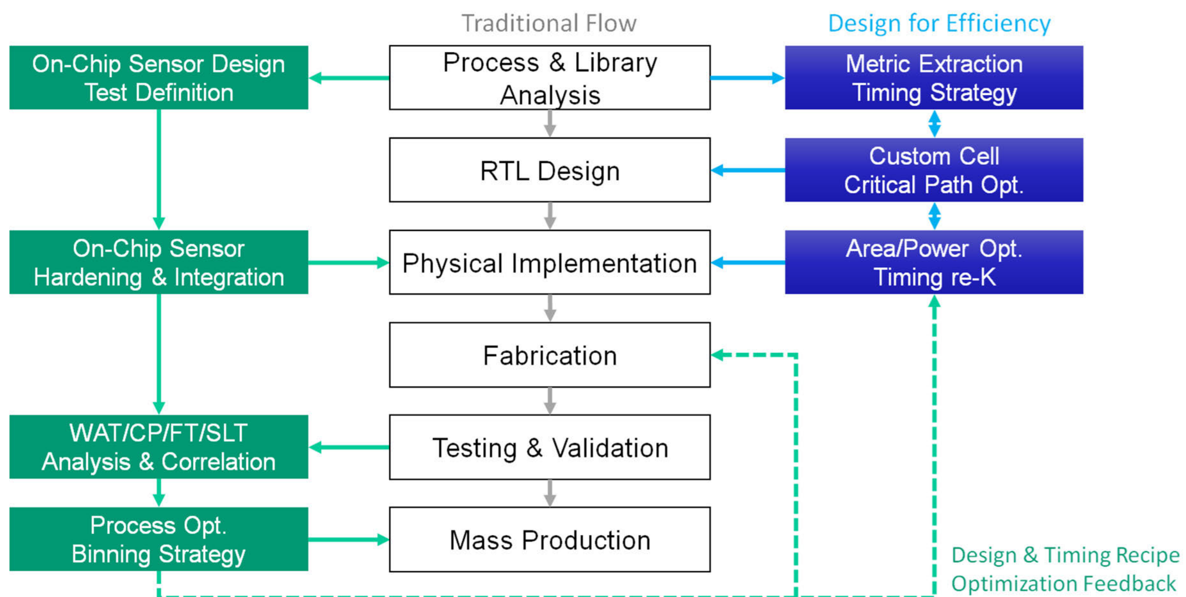

FIGURE 2. Flowchart of the actual implementation details of DTCO proposed in this study.

However, the acquisition and exchange of valuable test data often pose challenges and act as barriers to the overall advancement of industry technologies. Therefore, this study introduces an innovative virtual silicon technology that utilizes deep learning models to generate a significant volume of chip data rapidly and accurately. This technology accurately reflects the parameter distributions, defects, and features present in wafer manufacturing processes. Specifically, this study proposes a Generative Adversarial Network (GAN)-based approach that trains and encapsulates multidimensional WAT and CP test data using compact GAN models. As a result, highly realistic chip data with multidimensional features is generated. This technology plays a crucial role in optimizing chip design and improving the manufacturing process, leading to enhanced production efficiency, cost reduction, and improved product quality.

The primary contribution of this study is the introduction of an innovative virtual silicon technology that utilizes Generative Adversarial Networks (GANs) to rapidly and accurately generate large amounts of chip data, establishing a Design-Technology Co-Optimization (DTCO) platform and demonstrating its application in the platform.

The technology accurately reflects parameter distribution, defects, and characteristics present in the wafer manufacturing process. It plays a crucial role in optimizing chip design and improving manufacturing processes, thereby enhancing production efficiency, reducing costs, and improving product quality.

The remainder of the paper is organized as follows. Section II discusses the significance of quality control in semiconductor manufacturing and highlights the challenges in obtaining essential test data for process improvement. The section reviews existing literature and emphasizes the need for more advanced modeling and analysis methods, as well as explicitly stating research objectives. Section III details the GAN-based approach for generating virtual silicon data. It discusses the data transformation process, augmentation techniques, and the structure of the GAN model used. It also analyzes the complexity of the proposed method and elaborates on the practical applications and implications of the GAN-generated chip data. Section IV presents the experimental results of the generated silicon data using the proposed GAN model and its application in a collaborative optimization platform. The section includes visualizations and quantitative metrics to assess data quality. Finally, Section V summarizes the contributions of the study, emphasizing the potential of GANs and virtual silicon data in chip design, as well as the practical implications of the proposed approach in real-world semiconductor manufacturing. It acknowledges the challenges and the need for further research and development in this area.

II. BACKGROUND

In semiconductor manufacturing, quality control is crucial for cost savings and timely delivery [1], [2]. WAT is a critical step in semiconductor manufacturing, used to identify wafer defects, improve yield, and control costs [3], [4], [5]. However, WAT parameters are often high-dimensional, redundant, and challenging to obtain, hindering accurate yield prediction [6], [7]. CP identifies defective dies in wafers, reducing packaging and testing costs and providing yield information. Nevertheless, CP is time-consuming and expensive. Engineers use WAT-based yield prediction to save time and costs in lieu of CP [8], [9]. However, current key parameter identification methods have issues, including sacrificing selection model stability for single-parameter relationships and time efficiency for combined-parameter effects [10].

Xu et al. [6] proposed a hybrid feature selection method to identify key WAT parameters influencing wafer yield, which consists of two stages: filter selection and wrapper selection. In the filter selection stage, a minimum Redundancy Maximum Relevance (mRMR) filtering parameter pre-screening criterion based on Mutual Information (MI) is proposed. In the wrapper selection stage, a wrapped key parameter identification model based on Genetic Algorithm (GA) and Deep Belief Network (DBN) is designed. Fan et al. [11] proposed a new fault diagnosis framework that utilized a series of Machine Learning (ML) techniques for classification and data visualization. The study chose the Limited Random Synthetic Minority Oversampling Technique (LR-SMOTE) model, combined with vector-valued data augmentation and Principal Components Analysis (PCA) for the classification task. The findings suggest that this new ML solution shows great promise as a valuable tool for fault diagnosis of the WAT and CP processes in semiconductor manufacturing. Additionally, this study suggests that future research in this field should explore how to develop an appropriate version of Generative Adversarial Network (GAN) [12], [13] to augment the time series data of the process stations corresponding to faulty wafers.

Traditional models often simplify events by assuming Gaussian distributions, disregarding the fact that physical quantities in real chips frequently exhibit skewed-normal or log-normal distributions. Additionally, these models tend to overlook the interdependencies among vectors in high-dimensional spaces. Figure 3 demonstrates that even if each dimension's feature follows the distribution of the parent population when examined individually, the combined distribution in high-dimensional space may lose the interrelationship between them, analogous to rolling dice.

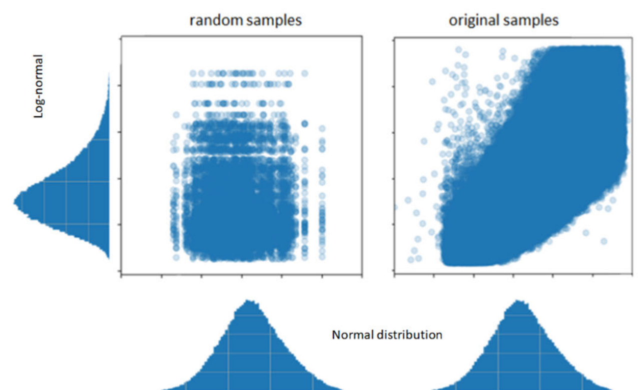

FIGURE 3. Interdependencies and correlation among features.

However, such simplifications may not accurately capture the various characteristics and interrelationships within the chip manufacturing process in real-world scenarios. To better simulate the behavior of real chips, it is crucial to consider features such as non-Gaussian distributions, skewed-normal distributions, or log-normal distributions while fully accounting for the interdependencies in high-dimensional space. Figure 4 illustrates the correct feature projection relationships: the probability distributions of features in each dimension follow the parent samples; the correlation distribution between any two features adheres to that of the parent samples; the correlation distribution among selected features in high-dimensional space conforms to the distribution observed in the parent sample.

Due to the multitude of process parameters involved in wafer manufacturing, the relationship between process parameters and wafer or chip-level test data becomes highly

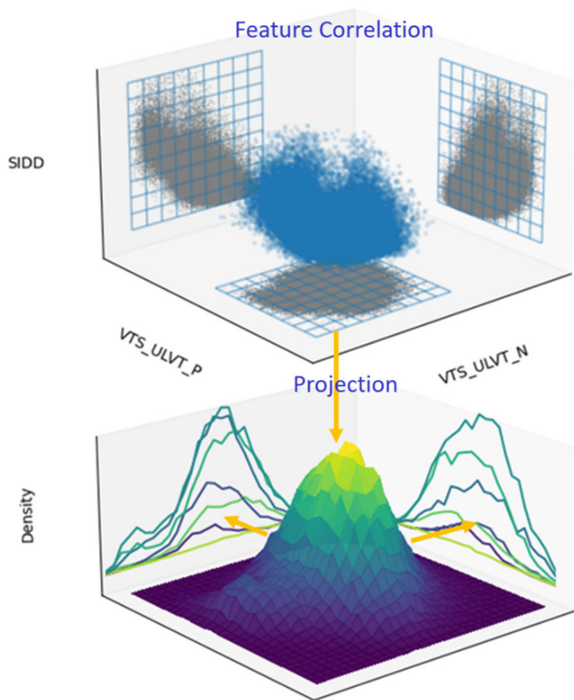

FIGURE 4. Feature projection.

intricate, making it challenging for traditional methods to effectively model and analyze. Figure 5 depicts that even with a comprehensive understanding of the distribution and interrelationship of chip-level features, the lack of wafer-level coordinate information results in the loss of characteristics related to the actual wafer fabrication uniformity. This is a prevalent issue in current simulation analysis modeling. In fact, the distribution of feature vectors in high-dimensional space lacks authenticity, leading to significant discrepancies between production data and simulated data.

This prejudice limits our understanding of the device modeling and affects the accuracy of simulations. To overcome this challenge, it is essential to develop more advanced modeling and analysis methods that fully consider the complexity of process parameters and the significance of wafer-level coordinate information. This will enable accurate simulation of chip characteristics and processes, providing more reliable production data and simulation results. It is of great importance for enhancing efficiency and quality control in the chip design industry.

Existing literature has investigated the utilization of deep learning models for simulating and predicting wafer manufacturing processes. Wang et al. [14] proposed a novel deep learning model called adaptive balancing generative adversarial network (AdaBalGAN) for the defective pattern recognition (DPR) of wafer maps with imbalanced data. Hu et al. [15] proposed a method for wafer defect detection based on a Deep Convolutional Generative Adversarial Network (DCGAN). This approach utilized a DCGAN to learn the distribution of defect images on wafers and employed the generated model for defect detection and classification. Furthermore, other research demonstrated the potential of Generative Adversarial Networks (GANs) for various other applications in the manufacturing domain [16], [17], [18], [19].

However, our research differs from existing literature in several aspects. Our research methodology primarily focuses on the application of GANs for generating virtual silicon data, which is crucial for optimizing chip design and improving the manufacturing process. We have successfully achieved the generation of highly realistic virtual silicon data and proposed a platform for the analysis and co-optimization of chip and wafer-level data based on GAN models. To better capture the variations in wafer-level processes, we have incorporated additional physical features into the construction of the training dataset. These improvements can enable more accurate simulations of the behavior in the chip manufacturing process and contribute to a better understanding of the interactions among chip characteristics. This will provide more precise simulation tools for chip design and manufacturing, aiding in optimizing chip performance and process control.

This study also provides a comprehensive analysis and optimization platform that can accurately simulate and predict various parameters and features in the chip manufacturing process, as shown in Figure 6. This will assist semiconductor fabs in better understanding and mastering process control, while improving production efficiency and product quality. Our approach not only exhibits high reliability but also promotes collaboration and knowledge sharing among interdisciplinary teams. Overall, our research brings new insights and solutions to the field of chip implementation.

Our research is making exciting breakthroughs in several key areas. Firstly, we achieve high-quality chip data generation by encapsulating millions of multi-dimensional data points into a compact GAN model, providing more efficient and accurate tools for chip manufacturing simulation and analysis. Secondly, we expand the scope of existing research, offering new ideas and directions while emphasizing the simplicity and practicality of the model, thereby providing feasible solutions and reliable tools for the field of chip manufacturing.

Our research outcomes have significant practical value as they provide clear guidelines for process recipes in semiconductor fabs, reduce chip-testing costs, and improve product quality. Moreover, they facilitate the co-optimization of chip design and process technology and foster cross-disciplinary collaboration. In summary, our research opens new avenues for simulation and analysis in chip co-optimization, laying a solid foundation for achieving more efficient and accurate methods in this field.

Based on the preceding background, the primary research objective of this study is to explore and implement application of GANs for generating wafer-level WAT and CP test data in chip manufacturing processes, with a focus on DTCO.

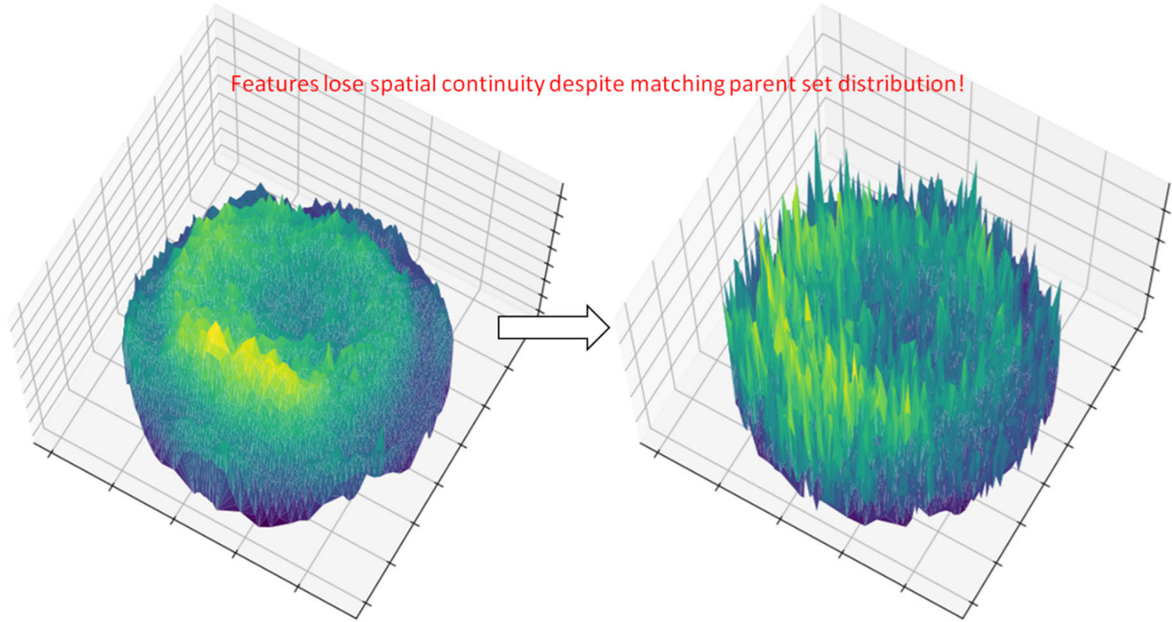

FIGURE 5. Loss of realism in the distribution of physical features of chips at the wafer-level.

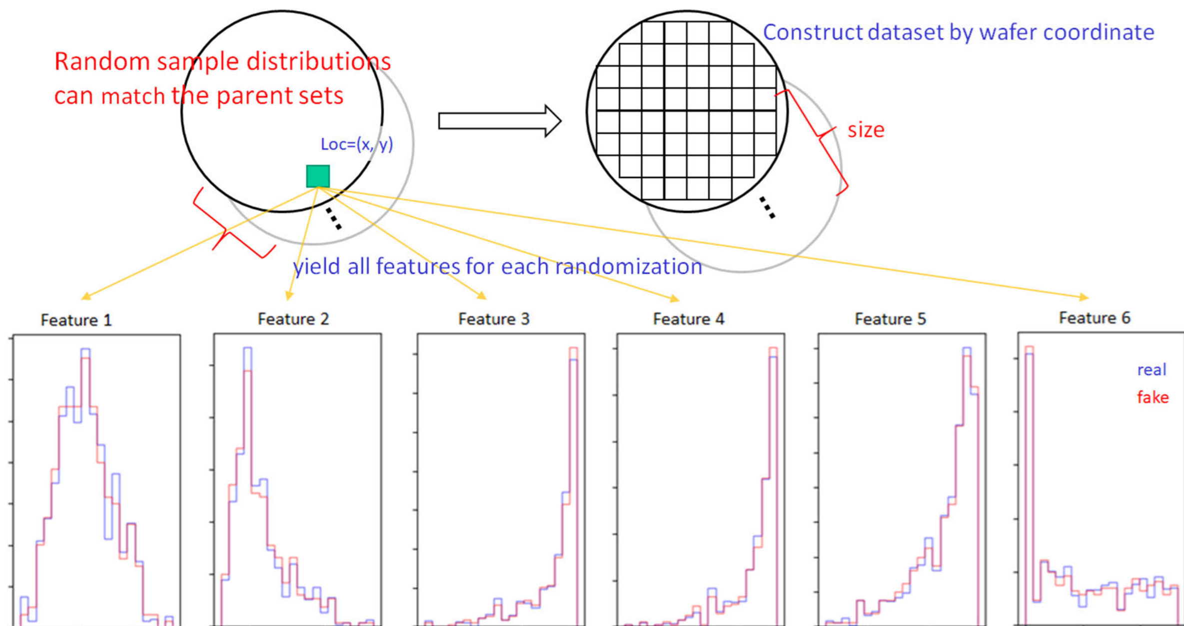

FIGURE 6. Probability distributions of generated data dimensions matched to the original population samples.

We aim to establish a working model for advancing DTCO and illustrate its potential using a fundamental GAN model. Through this model, we intend to generate virtual silicon data to realize several tangible advantages of DTCO. This will indirectly highlight the potential within the Electronic Design Automation (EDA) field for developing models, exploring algorithms, and creating tools through collaborations with both industry and academia.

III. SYSTEM ARCHITECTURE

This study employs a Generative Adversarial Network (GAN) model to capture the uniformity characteristics of defects and parameters in the wafer manufacturing process using a large volume of multi-dimensional data. GANs consist of two neural networks, a generator, and a discriminator, working in a competitive manner. The generator generates data, while the discriminator evaluates the generated data's authenticity.

The two networks are trained iteratively, with the generator striving to create data that's indistinguishable from real data, and the discriminator trying to tell the real from the fake. This back-and-forth process continues until the generated data closely matches real data. Furthermore, data augmentation techniques are used in conjunction with GANs. Data augmentation is a method for increasing the size of the dataset by applying various transformations to the original data. For example, it can involve rotating, cropping, or otherwise altering the data to create additional variations for training. These techniques are crucial for improving the efficiency of our approach in chip and wafer test data modeling. Through the implementation of GANs and data augmentation, we enhance our capability to generate realistic and diverse data. This approach contributes to the optimization of chip design and manufacturing processes, enabling the development of more accurate and efficient simulation tools. As such, it plays a significant role in our research. Initially, the proposed platform transforms the original multi-dimensional data into two-dimensional images, and set multiple feature dimensions (parameter C), as shown in Figure 7. The size of parameter C is correlated with the network size, and the training time exhibits non-linear growth. Based on computations performed on a personal computer CPU, the proposed platform selects the feature dimensionality C to be between 10 and 18.

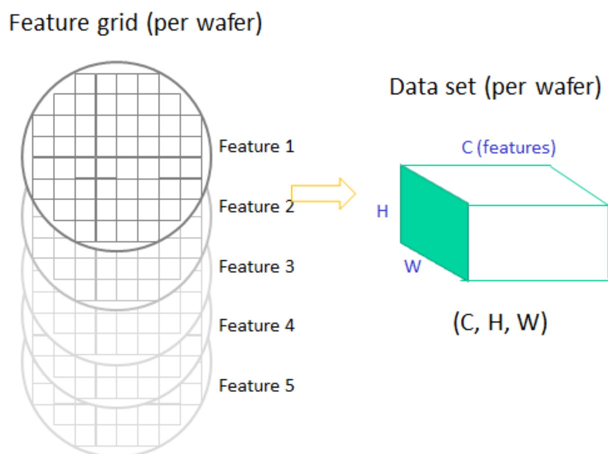

FIGURE 7. **Transformation and integration of wafer-level multidimensional training dataset.**

We attempted training on an ordinary laptop (Intel Core i7-1255U, 1.70GHz, 32GB RAM). With a feature size limit set between 10 to 18 dimensions, the training time was 8 hours, resulting in a model of decent quality. To further increase the feature size, we explored different generative model architectures, such as the Wasserstein Generative Adversarial Network (WGAN). In comparison, WGAN showed roughly seven times faster convergence and potential for further improvements in similarity compared to this study.

In this study, we further augments the training dataset using small angle rotation transformations to simulate the occurrence of rotational defects and process parameters distributions in the wafer manufacturing process, as shown in Figure 8. This augmentation technique enables us to accurately capture the key features in the wafer manufacturing process, thereby improving the training effectiveness of the model.

This study utilizes a Convolutional Neural Network (CNN) to construct the Generative Adversarial Network (GAN) model, as shown in Figure 9. The GAN model is responsible for generating chip data with various process features while incorporating potential defects. The generator component of the model consists of multiple convolutional layers and Tanh activation layers to generate wafer images. Simultaneously, the discriminator component also includes multiple convolutional layers and Sigmoid activation layers to distinguish between real and generated data. These design components work together to achieve the goal of generating high-quality silicon data.

This study presents algorithms 1 and 2 to implement the proposed GAN model for chip and wafer test data modeling. The algorithms define the following functions: $f_{VSD}(\ )$ for Virtual Silicon Data preprocessing and $f_{GAN}(\ )$ for GAN Training aimed at Virtual Silicon Generation.

GAN training consists of optimizing two components: the discriminator's parameters, which are adjusted to maximize classification accuracy, and the generator's parameters, which are tuned to thoroughly deceive the discriminator. The generator G creates images from random noise. In other words, it learns a mapping from a random noise vector z to an image y, G: $z \rightarrow y$. Simultaneously, the discriminator D learns a mapping from an image x to some values between 0 to 1, which indicates the probability that the input comes from the real data distribution. The training process evaluates cost through a value function, denoted as $V(G, D)$, which considers both the generator and the discriminator. The training process can be expressed as follows:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x\sim pdata(x)}[\log D(x)] + E_{z\sim pz(z)}[\log(1 - D(G(z)))]$$

During the model training process, this study employs the gradient descent optimization algorithm to minimize the difference between the generated chips and real chips. To enhance the stability of the model, this study utilizes techniques such as batch normalization and the LeakyReLU activation function. Through several hundred iterations, the GAN model is capable of generating highly realistic silicon data, including the chip's position on the wafer, the uniformity of physical features at the wafer level, and the defects present in the chip manufacturing process. The training results provide a reliable data basis for the simulation and analysis of the chip implementation and optimization process.

Because semiconductor chip test data is not similar to images of people or scenes in real life, which can intuitively represent quality visually, this study utilizes Jensen-Shannon

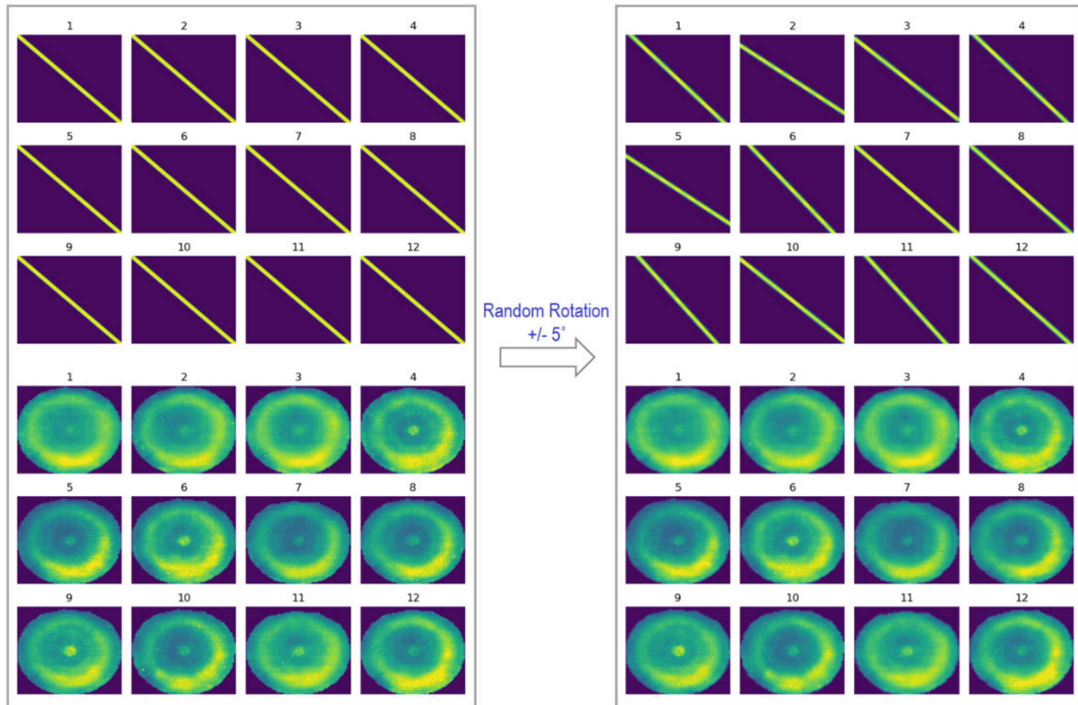

FIGURE 8. Augmentation of training dataset through small angle rotation transformations.

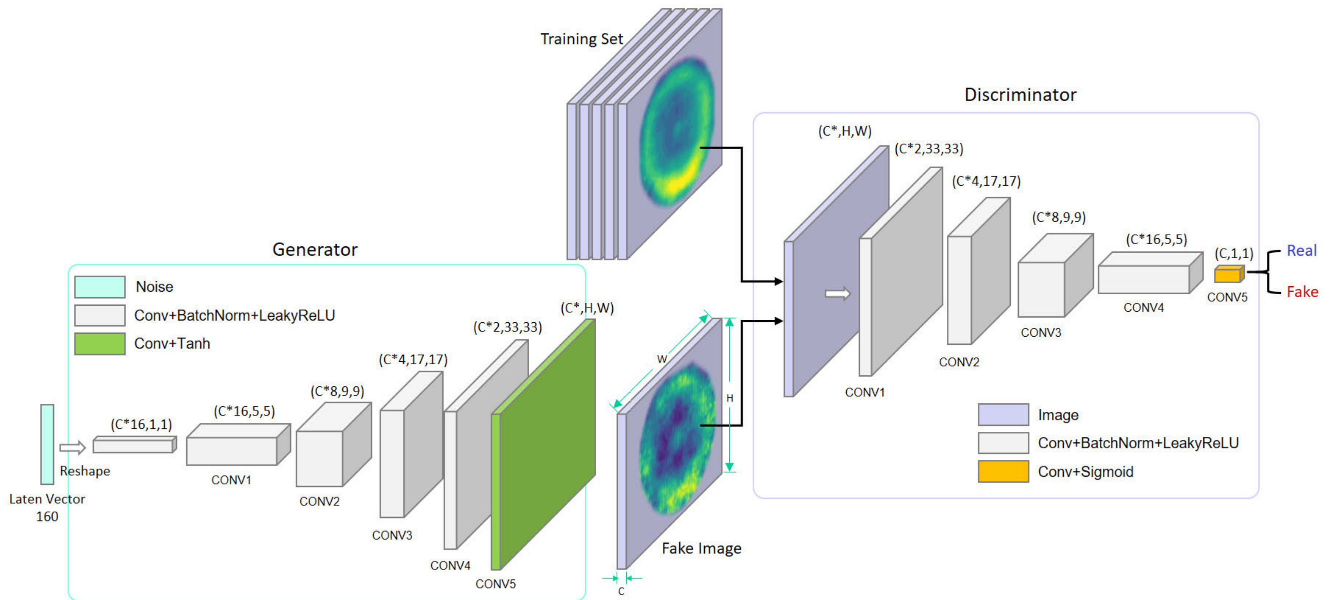

FIGURE 9. The architecture of the proposed GAN model.

Divergence to demonstrate the similarity between individual dimensions at the lowest level and the parent sample. Additionally, it combines spatial distributions of two or more dimensions to illustrate the quality of data generation.

The most challenging aspect of the framework in this study is the adjustment of training hyperparameters. A too large learning rate can lead to model divergence, while a too small learning rate results in very slow training.

Algorithm 1 VSDataset Preprocessing

1. Replace any NaN values in the data with zero.
2. Calculate the per-channel minimum and maximum values in the data for normalization.
3. Normalize the data to the range [0, 255] and convert it to uint8.
4. Convert the mask to a boolean tensor.
5. Define a transformation pipeline with *ToTensor*, *RandomRolation*, and *Normalize* operations.

Algorithm 2 GAN Training for Virtual Silicon Generation

Input: *Generator (G)* and *Discriminator (D)* models.
Output: Trained *Generator* (G) and *Discriminator* (D) models.

1. Initialize *Generator* (G) and *Discriminator* (D) models.
2. Define Binary Cross-Entropy Loss Critcrion.
3. Initialize Optimizers for Generator and Discriminator.
4. Prepare chip data (ds) and create a DataLoader *(VSDataset)* for batching.
5. **Training Loop:**
6. For each epoch in range (num epochs):
7. For each batch (images) in daiaioader:
8. Generate random noise (noises) for fake chip data.
9. Train the Discriminator:
10. Compute discriminator loss for real and fake data.
11. Backpropagate and update the Discriminator's parameters.
12. Train the Generator:
13. Generate lake chip data using the Generator.
14. Compute generator loss.
15. Backpropagate and update the Generator's parameters.

Therefore, in practice, it is advisable to employ a stepwise learning rate (lr) adjustment, such as dynamically changing the learning rate every 100 epochs as follows: lr: = gamma * lr, where gamma = 0.8, and the initial lr is set to 0.001. In practical applications, the above suggestions are by no means the ultimate solution. There is always room for further optimization.

In this study, we also explore the option to employ the conventional GAN model and optimize our loss using the Adaptive Moment Estimation (Adam) optimization algorithm in conjunction with Binary Cross Entropy (BCE). This methodology can also be adapted to utilize WGAN. The adaptation involves eliminating the BatchNorm layer from the Discriminator and replacing the BCE loss function with Wasserstein loss function. This modification results in a nearly sevenfold increase in the model's convergence speed, with an average Jensen-Shannon Divergence similarity exceeding 99%. It is important to clarify that this study does not intend to emphasize neural network algorithms. Instead, our primary goal is to showcase the potential of generating virtual silicon data using a straightforward GAN. We offer a practical demonstration of a working model applicable in various DTCO scenarios, utilizing the generated virtual silicon data for the development of viable EDA tools.

In reality, the industry recognizes that DTCO is crucial for achieving real chip efficiency and optimizing productivity, encompassing factors like yield, cost, and overall competitiveness. However, acquiring real data is a formidable challenge. These datasets contain confidential information that significantly impacts business competitiveness, leading companies with access to such data to be reluctant to share it. Furthermore, these datasets are often massive and not easily transferable, hindering the establishment of practical data exchange and working models. This study introduces a feasible working model that comprehensively addresses this issue, covering foundational aspects such as the implementation flow, sensor integration, data analysis, and the DTCO platform, all the way to practical applications. Through the use of generative neural network models, large volumes of data can be compressed, concealed, scaled, and normalized to protect sensitive information. This simple generative model facilitates data interchange across networks, fostering academic research and continuous improvement. Furthermore, it opens up numerous new opportunities for the development of EDA algorithms and tools within the industry. It's important to note that this methodology is not limited to virtual silicon data but provides a potential working model for DTCO.

IV. EXPERIMENTAL RESULTS AND COLLABORATIVE OPTIMIZATION PLATFORM

This section presents the generated chip data using the proposed GAN model of this study and conduct a detailed analysis, while establishing a Design-Technology Co-Optimization (DTCO) platform. The dataset used consists of approximately 12 million chip data points, excluding 3σ outliers and missing chip data. Notably, chips with systematic defects are deliberately retained in the training set for the GAN model. This study utilizes both data visualizations and quantitative metrics to evaluate the quality of the generated silicon data. For instance, this study uses Jensen-Shannon Divergence to compare the similarity of probability distributions between the generated data and real chip data. Additionally, this study leverages the Kernel Density Estimation (KDE) metric to quantify the numerical differences between probability distributions of different features. These evaluation methods ensure a reliable and accurate assessment of the quality of the generated silicon data.

To protect the confidentiality of chip technology and wafer process data, all charts and figures in the research results have been uniformly normalized, limiting the numerical range between 0 and 1.

The experimental results demonstrate that the scatter plots between the features of the generated silicon data by the GAN model and the real chip data exhibit a high degree of similarity, effectively capturing the process adjustment and variability in the early stages, as shown in Figure 10. This figure illustrates the similarity between the generated four-dimensional feature vectors and the distribution of real data. The left side displays CP features

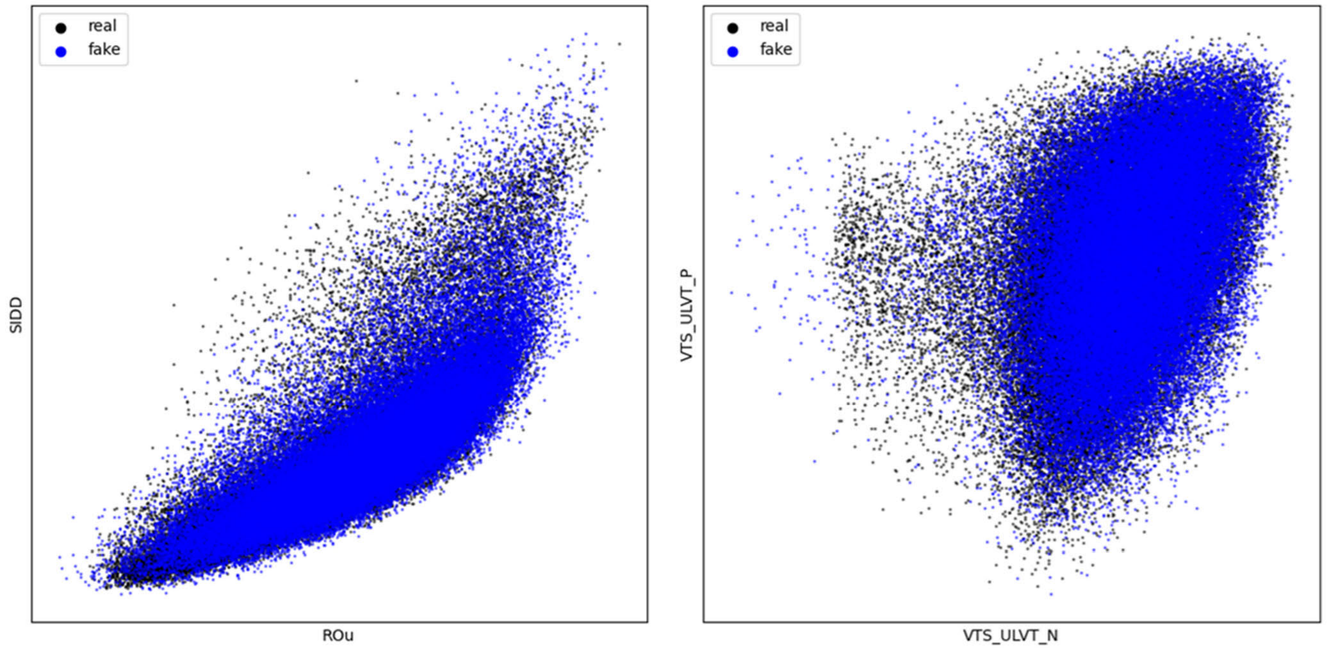

FIGURE 10. Scatter plots of features for generated data and real data.

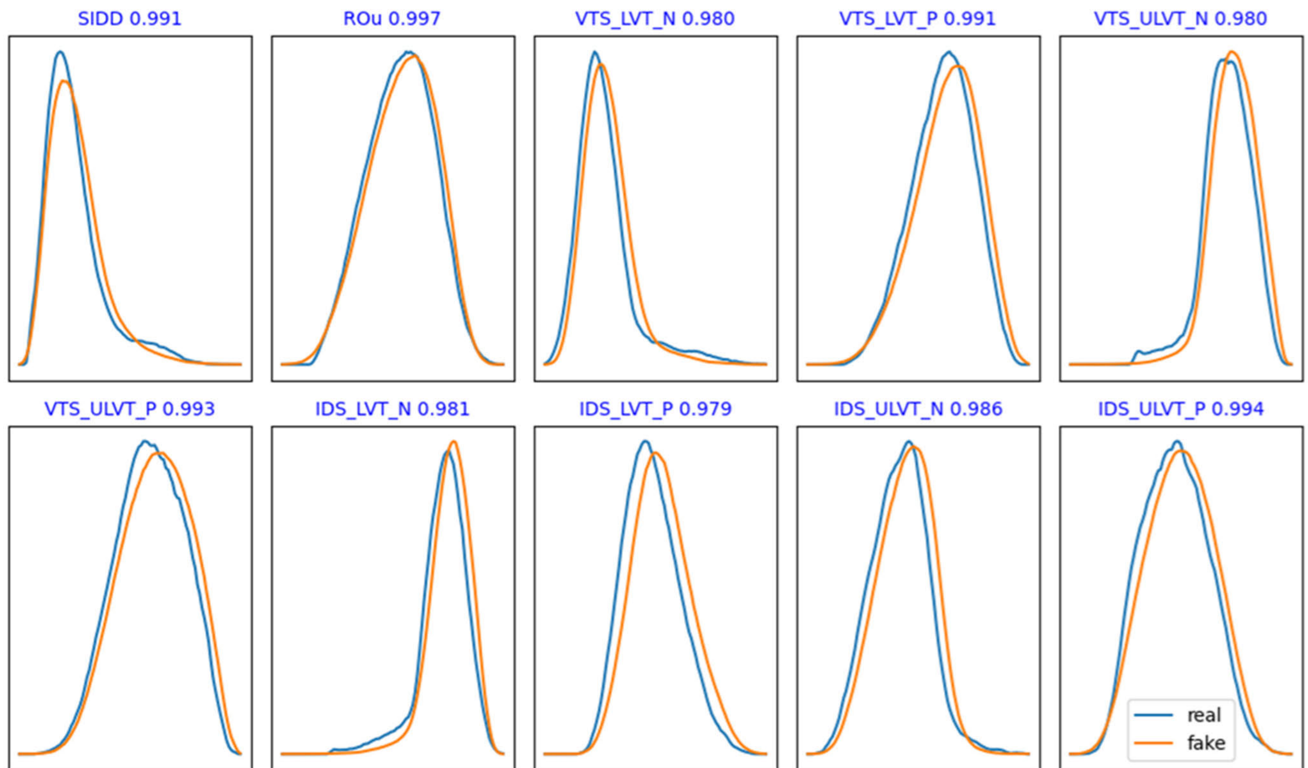

FIGURE 11. Probability distributions of features for generated data and real data.

(X: Rou represents chip speed, Y: SIDD represents leakage current), while the right side shows WAT features (X: VTS_ULVT_N represents NMOS gate threshold voltage, Y: VTS_ULVT_P represents PMOS gate threshold voltage). Moreover, further analysis using the Jensen-Shannon Divergence index reveals that the probability distributions of the generated silicon data closely align with the characteristics of the real data, ranging from 0.98 to 1.0 across different

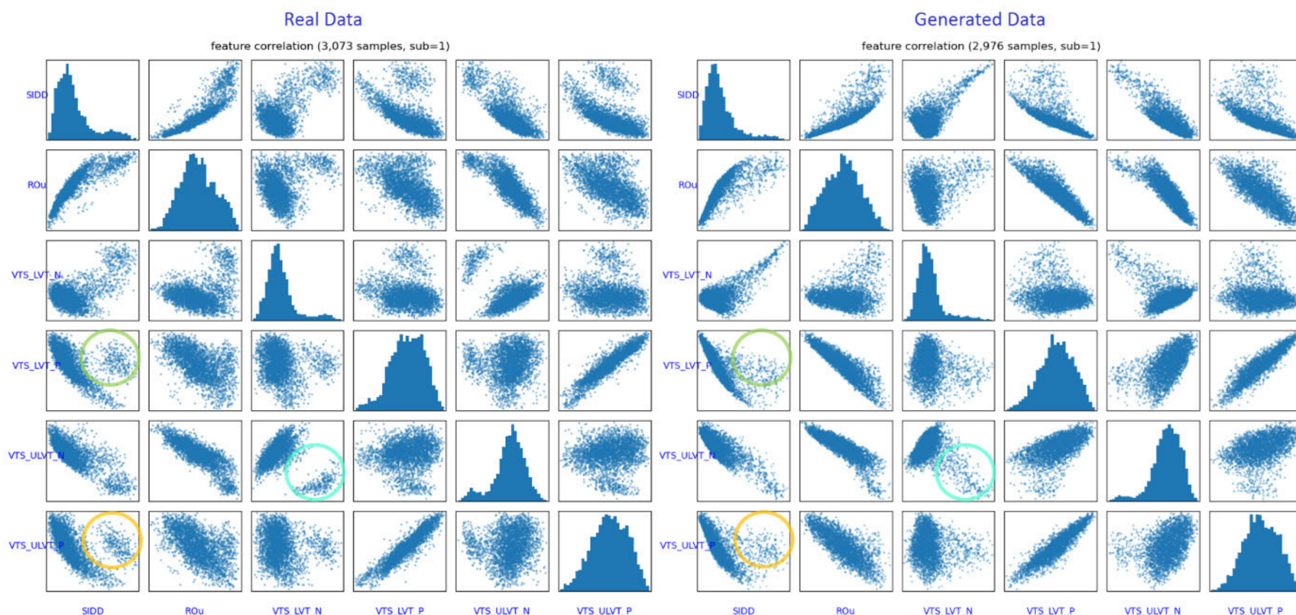

FIGURE 12. Preservation of correlations between the combination of generated data in high-dimensional space and the original real data.

FIGURE 13. Feature uniformity of 15 generated wafer data.

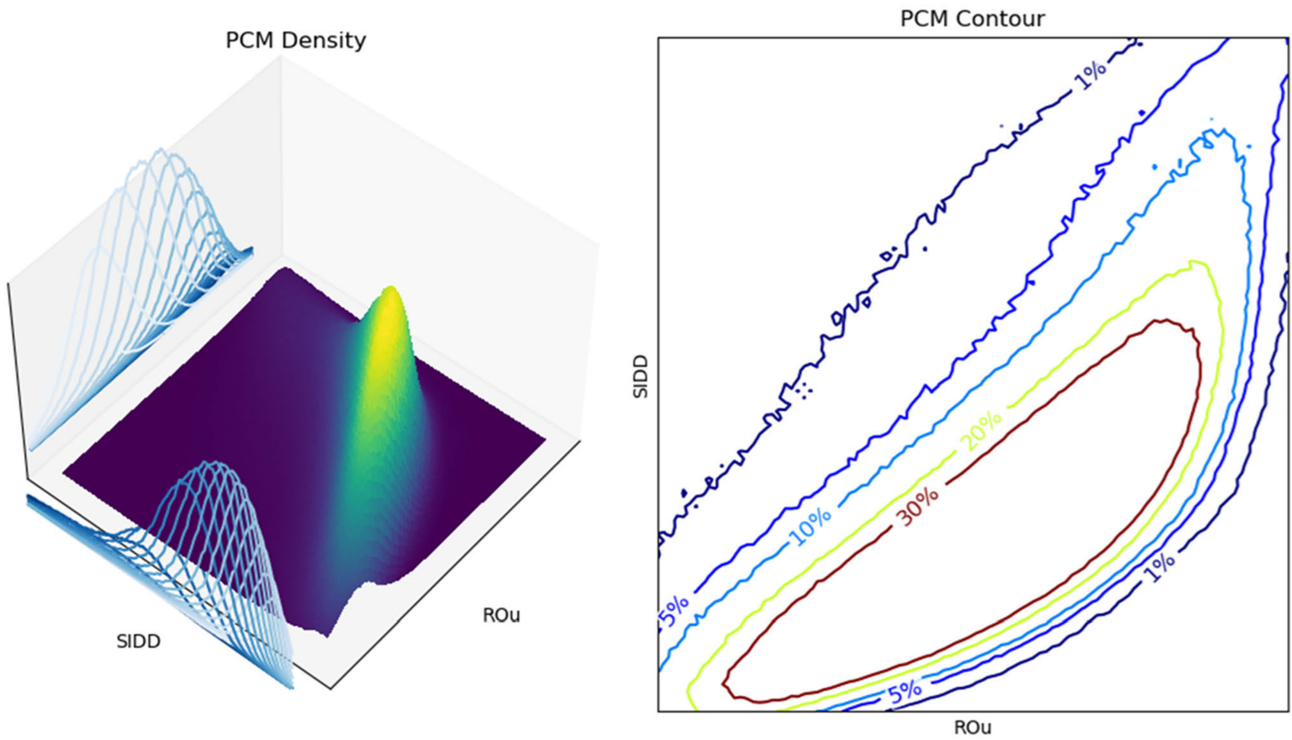

FIGURE 14. Probability density distribution of multidimensional features.

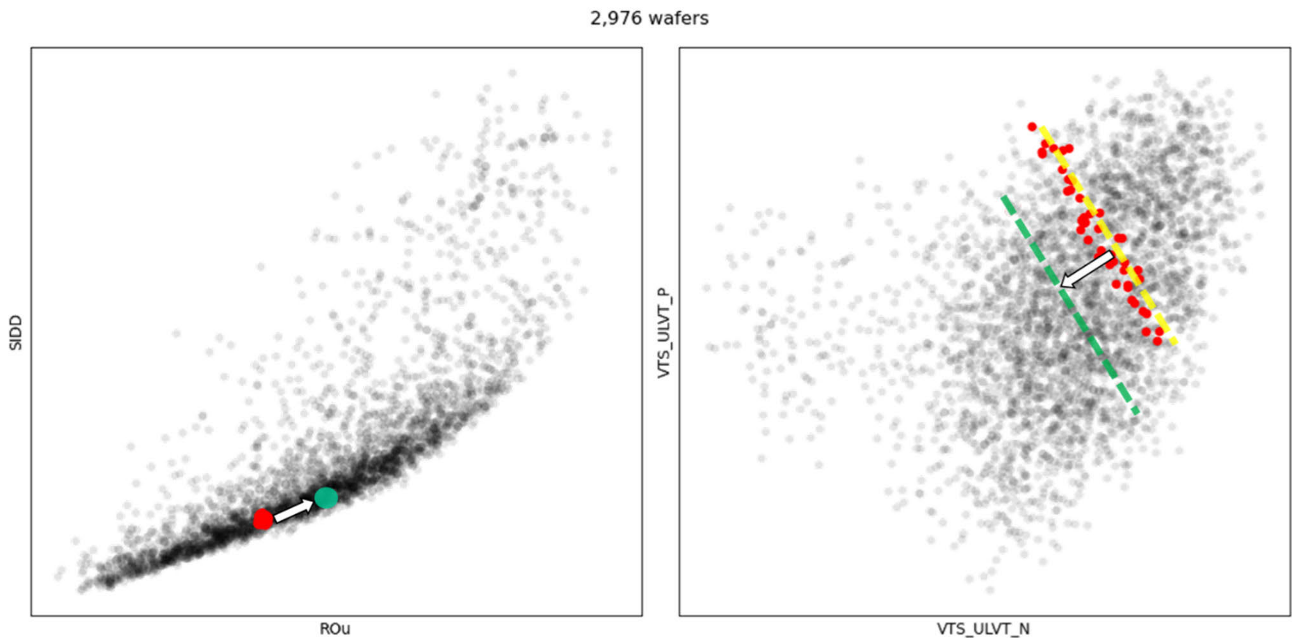

FIGURE 15. Cross-probing between WAT and CP.

dimensions, as shown in Figure 11. Furthermore, the combination of generated data in high-dimensional space still preserves the correlations of the original real data, as shown in Figure 12.

To enhance the quality of the generated model, we can selectively omit some parameters that are not important in actual production to avoid the model learning abnormal values during the early stage of process parameter adjustment.

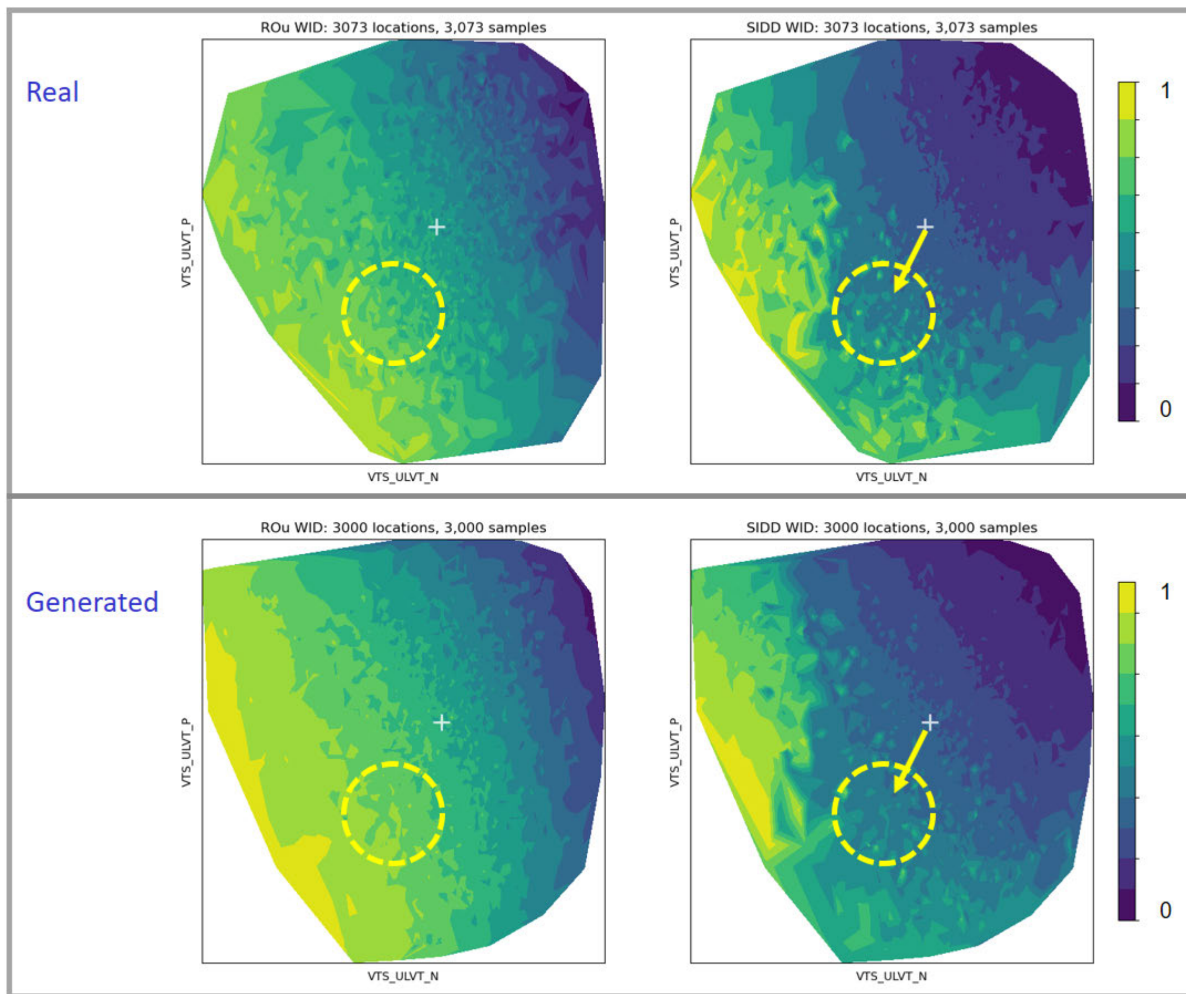

FIGURE 16. **Process recipe optimization.**

GAN models are more effective in learning the non-linear relationships and feature representations in chip manufacturing processes, capturing fine details and better fitting the distribution of real data compared to traditional methods that analyze real data. Further analysis demonstrates that GAN models can learn and capture the parameter distribution and uniformity details in the chip manufacturing process, as shown in Figure 13. This information on uniformity represents non-random features on wafers, such as similar distorted wafer surfaces observed in lots produced from various batches.

Figure 14 shows the probability density distribution of multidimensional features. It illustrates the compromise space between yield and design margin based on the generated large dataset of chip data, providing specific guidance for future design recipes and capacity optimization. The proposed GAN model of this study demonstrates good stability and generalization performance across different training and testing sets.

The cross-dimensional data correlations generated by the GAN model offer exciting applications for Design-Technology Co-Optimization (DTCO). We observe correlations between wafer-level process parameters, such as the threshold voltage (VT) of N-MOS and P-MOS, and chip-level performance (Fmax) and leakage current (SIDD), as shown in Figure 15. The variability of the NP process recipe exhibits a negative slope coefficient and intercept. Through the analysis of cross-dimensional correlations and data interactions, we can more effectively identify trade-offs and optimization directions between target energy efficiency and process recipes.

Figure 16 presents another application of high-dimensional feature profiling. By conducting cross-dimensional and correlation analysis, such as utilizing wafer-level process parameters as a reference, we can efficiently provide targeted process adjustment strategies and guidelines to enhance chip performance (e.g., an averaged ring-oscillator frequency; ROu) while preserving the same characteristic (e.g., SIDD). This enables us to optimize the process while preserving certain desired qualities.

Figure 15 can be likened to a dynamic language translation expert and bridge, facilitating mutual understanding

between experts on the left side representing the design house and those on the right side representing the semiconductor fab. This mutual understanding is achieved through real-time cross-probing displayed on the user interface, enabling both sides to comprehend each other's objectives and directions. Similarly, Figure 16 can be considered a static content translation expert and bridge. It translates the XY plane coordinates using the parameters of the semiconductor fab and the Z-axis using the parameters of the design house. This is achieved through high-dimensional contours that present optimization objectives and specific directions, which are clear to both parties. For example, a typical designer may desire a chip with reasonable speed and power efficiency. Traditionally, this would involve extensive analysis of real chip efficiency using split wafers, followed by iterative adjustments of the manufacturing parameters with the semiconductor fab to find the best process window. Nevertheless, by integrating data across different dimensions and projecting data into higher dimensions, we can establish a model much like a translation expert or bridge that enables experts from different domains to immediately and accurately identify the process window positions that suit each other, expediting the convergence process. Figure 16 also demonstrates that the specific best process window and optimization directions provided by virtual data match the original data. This is highly exciting.

In order to evaluate the practical applications of the generated chip data in DTCO, a comprehensive series of tests was conducted. These tests encompassed various critical aspects, such as the distribution and uniformity of process parameters, detection of wafer surface defects, trade-offs between process parameters and design margins, chip performance prediction, optimization of productivity and binning strategies, among others. The experimental outcomes confirmed the promising potential of our GAN-generated chip data in facilitating the development of novel Electronic Design Automation (EDA) tools for DTCO-related domains.

V. CONCLUSION

Defect detection and classification are critical steps in any semiconductor manufacturing process, involving a significant investment of time, materials, and labor costs. The primary contribution of this study is the introduction of an innovative virtual silicon technology that utilizes deep learning models to rapidly and accurately generate a substantial volume of chip data. This establishes a Design-Technology Co-Optimization (DTCO) platform and demonstrates its application in the platform. The technology precisely reflects the parameter distributions, defects, and characteristics present in the wafer manufacturing processes. Specifically, the study introduces a Generative Adversarial Network (GAN)-based approach to train and encapsulate multidimensional WAT and CP test data, resulting in the generation of highly realistic chip data with multidimensional features. The research outcomes demonstrate that the proposed approach effectively assists in chip design and wafer production for product optimization and process improvement, resulting in enhanced production efficiency, reduced costs, and improved product quality. This paves the way for novel approaches in simulating and analyzing chip co-optimization.

However, GAN also faces several challenges, such as selecting the appropriate architectures for the generator and discriminator, handling high-dimensional and complex data, and requiring significant time and computational resources. To better transform data into trainable models, we adopt a method of converting multidimensional data into two-dimensional images with multiple feature channels. Additionally, we simulate rotation defects and variations in process parameters that may exist in wafer manufacturing during the training process.

In summary, we believe that generated models, including but not limited to GAN, hold immense potential in the development of new Electronic Design Automation (EDA) tools. By utilizing these generated models, clear guidelines for design and process recipes can be provided, leading to reduced chip testing costs and improved product quality. Furthermore, it fosters collaborative optimization between chip design and process technology across different domains. However, achieving this potential requires further research and development to overcome existing challenges and limitations. We look forward to more researchers addressing this issue and proposing innovative solutions, laying a solid foundation for achieving more efficient and accurate chip co-optimization methods.

REFERENCES

[1] Y. Zheng, D. Ling, Y.-W. Wang, S.-S. Jang, and B. Tao, "Model quality evaluation in semiconductor manufacturing process with EWMA run-to-run control," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 8–16, Feb. 2017.

[2] J. Ruth and R. Berndt, "Quality control for ultrafiltration of ultrapure water production for high end semiconductor manufacturing," in *Proc. 27th Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, Saratoga Springs, NY, USA, May 2016, pp. 16–22.

[3] G. S. May and S. M. Sze, *Fundamentals of Semiconductor Fabrication*. Hoboken, NJ, USA: Wiley, 2003.

[4] J. N. D. Gupta, R. Ruiz, J. W. Fowler, and S. J. Mason, "Operational planning and control of semiconductor wafer production," *Prod. Planning Control*, vol. 17, no. 7, pp. 639–647, Oct. 2006.

[5] D.-H. Lee, J.-K. Yang, C.-H. Lee, and K.-J. Kim, "A data-driven approach to selection of critical process steps in the semiconductor manufacturing process considering missing and imbalanced data," *J. Manuf. Syst.*, vol. 52, pp. 146–156, Jul. 2019.

[6] H. Xu, J. Zhang, Y. Lv, and P. Zheng, "Hybrid feature selection for wafer acceptance test parameters in semiconductor manufacturing," *IEEE Access*, vol. 8, pp. 17320–17330, 2020.

[7] H.-C. Hong and L.-Y. Lin, "Accurate and fast on-wafer test circuitry for device array characterization in wafer acceptance test," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 9, pp. 3467–3479, Sep. 2019.

[8] F. Tan, T. Pan, Z. Li, and S. Chen, "Survey on run-to-run control algorithms in high-mix semiconductor manufacturing processes," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1435–1444, Dec. 2015.

[9] J. Moyne, J. Samantaray, and M. Armacost, "Big data capabilities applied to semiconductor manufacturing advanced process control," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 4, pp. 283–291, Nov. 2016.

[10] W. Zhang, P. Duan, L. T. Yang, F. Xia, Z. Li, Q. Lu, W. Gong, and S. Yang, "Resource requests prediction in the cloud computing environment with a deep belief network," *Softw., Pract. Exper.*, vol. 47, no. 3, pp. 473–488, Mar. 2017.

[11] S. S. Fan, C.-W. Cheng, and D.-M. Tsai, "Fault diagnosis of wafer acceptance test and chip probing between front-end-of-line and back-end-of-line processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3068–3082, Oct. 2022.

[12] K. Yan, J. Su, J. Huang, and Y. Mo, "Chiller fault diagnosis based on VAE-enabled generative adversarial networks," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 1, pp. 387–395, Jan. 2022.

[13] D.-M. Tsai, S. S. Fan, and Y.-H. Chou, "Auto-annotated deep segmentation for surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[14] J. Wang, Z. Yang, J. Zhang, Q. Zhang, and W. K. Chien, "AdaBalGAN: An improved generative adversarial network with imbalanced learning for wafer defective pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 310–319, Aug. 2019.

[15] G. Hu, J. Huang, Q. Wang, J. Li, Z. Xu, and X. Huang, "Unsupervised fabric defect detection based on a deep convolutional generative adversarial network," *Textile Res. J.*, vol. 90, nos. 3–4, pp. 247–270, Feb. 2020.

[16] A. Kusiak, "Convolutional and generative adversarial neural networks in manufacturing," *Int. J. Prod. Res.*, vol. 58, no. 5, pp. 1594–1604, Mar. 2020.

[17] R. Singh, R. Garg, N. S. Patel, and M. W. Braun, "Generative adversarial networks for synthetic defect generation in assembly and test manufacturing," in *Proc. 31st Annu. SEMI Adv. Semiconductor Manuf. Conf. (ASMC)*, Saratoga Springs, NY, USA, Aug. 2020, pp. 1–5.

[18] N. Hertlein, P. R. Buskohl, A. Gillman, K. Vemaganti, and S. Anand, "Generative adversarial network for early-stage design flexibility in topology optimization for additive manufacturing," *J. Manuf. Syst.*, vol. 59, pp. 675–685, Apr. 2021.

[19] J. Qin, F. Gao, Z. Wang, D. C. Wong, Z. Zhao, S. D. Relton, and H. Fang, "A novel temporal generative adversarial network for electrocardiography anomaly detection," *Artif. Intell. Med.*, vol. 136, Feb. 2023, Art. no. 102489.

SHIH-NUNG CHEN received the B.S. and Ph.D. degrees in computer science and information engineering from Tamkang University, Taipei, Taiwan, in 1995 and 2001, respectively. Since August 2001, he has been an Associate Professor and the Chairperson of the Department of Information Communication, Asia University, Taichung, Taiwan. His current research interests include computer simulation, distributed systems, bioinformatics, image processing, and artificial intelligence.

SHI-HAO CHEN received the Ph.D. degree in computer science from National Tsing Hua University, Hsinchu, Taiwan, in 2012. He is currently the CEO of DigWise Technology, a solution provider for best energy efficiency, and the CEO of DipSci Technology, a solution provider that provides on-chip process monitoring and machine learning technologies to improve productivity. His current research interests include machine-learning techniques for EDA, low-power, and process-design co-optimization methodologies.