Acceleration Of Chip Verification Process and Defect Prediction Based on Artificial Intelligence

Yizhe Lu *

School of Electronic Engineering, Xi'an University of Posts & Telecommunications, Xi'an, 710000, China

* Corresponding Author Email: 23072053@stu.xupt.edu.cn

Abstract. With the popularization of advanced manufacturing processes (3nm/2nm) and heterogeneous integration technologies, chip verification is facing challenges such as sensitivity to process deviations, complex physical effects, and a sharp increase in data volume. Traditional verification methods based on experience, such as test case writing and full-chip simulation, are inefficient and prone to missed detections. This article systematically reviews the application of artificial intelligence technology in accelerating the chip verification process and predicting defects. This paper focuses on analyzing how AI can address issues such as missed detections of process-related defects in microelectronics verification, bottlenecks in physical effect simulation, and multi-source data fusion of design parameters, process data, and test results. By combining accessible tools and open-source datasets in the microelectronics field, it explores the feasibility of technology implementation. Research shows that AI can increase the accuracy of predicting process-related defects to over 85% and shorten the simulation cycle by 40%. Future research can enhance interpretability and maintain the ability of efficient reasoning from the hybrid model that integrates symbolic reasoning and neural networks. Meanwhile, the collaborative mechanism of cross-scale verification tasks and the lightweight deployment of AI will become the key technical paths to promote the intelligence of EDA tools and the development of the open-source chip ecosystem.

Keywords: Chip verification; Artificial intelligence; Process-related defects; multi-source data fusion; simulation cycle optimization.

1. Introduction

Process deviations are highly sensitive. In processes below 7nm, lithography errors with a line width deviation of ±5nm and doping fluctuations with a threshold voltage drift of ±10mV can lead to chip performance dispersion, such as a frequency deviation of ±15%. However, traditional simulation can only cover the nominal process scenarios and fail to detect process-related defects. The physical effects are extremely complex. Physical phenomena such as IR voltage drop, power network voltage drop, and electromigration, like metal wire atom migration, require multi-physics field coupling simulation, and the calculation time consumption accounts for more than 30% of the verification cycle. Data fragmentation is severe. Design parameters such as transistor size, process data such as photoresist thickness, and test results are scattered across different tools, making it difficult to correlate and analyze them. With the rapid development of generative AI and large language models, many breakthroughs have emerged in the field of chip verification. In terms of defect prediction, Google Research has proposed a defect propagation prediction model based on a Transformer, which achieves cross-layer defect correlation analysis through a self-attention mechanism, reducing the missed detection rate compared to traditional GNN models [1].In the field of physical simulation, NVIDIA's PhysX-AI framework, for the first time, combines neural networks with finite element analysis, achieving a three-order-of-magnitude increase in the prediction speed of electromigration effects [2].In terms of multi-source data fusion, the AutoV verification platform developed by Samsung Electronics in collaboration with the University of California, Berkeley, integrates design, process, and test data silos through federated learning technology, shortening the verification cycle while ensuring privacy and security [3]. The open-source ecosystem is becoming an important driving force. The AI-Driven Verification Toolkit released by the OpenEDA Alliance integrates PyTorch-based process parameter optimization algorithms and has supported the rapid verification of RISC-V

processor cores [4]. These advancements indicate that AI is evolving from a single tool optimization to a systematic verification framework, providing a new paradigm for addressing the verification challenges of processes below 3nm.

MIT used graph neural networks to predict the impact of process deviations on transistor performance [5], and TSMC optimized the lithography process parameters using reinforcement learning methods [6]. AI can optimize the verification process. Through research, the advantages of AI in predicting process-related defects, accelerating physical effect simulation, and verifying multi-source data fusion, such as the linkage of design, process, and testing, have been discovered. Therefore, this paper systematically reviews the application of AI in the key links of chip verification, specifically including defect prediction under process deviation, simulation acceleration of physical effects, and fusion verification of design-process-test data. It comprehensively reviews the application of AI in verification, providing a cross-disciplinary research path for the microelectronics profession.

2. Application of Graph Neural Networks in Chip Defect Prediction

2.1. Application

The core aspect of graph neural networks in chip defect prediction is that GNNS model the microstructure of chips through node, transistor/interconnect to edge physical, and connection/process parameter networks, and uses graph convolution operations to capture the propagation law of process deviations at the device layer. MIT research shows that by using graph neural networks to model the coupling effect of photolithography and etching processes and learning nonlinear defect generation patterns, potential failure points that were overlooked in the simplified assumptions of physical simulations can be identified. The key to his technology lies in constructing multi-process coupled graph structure features and leveraging the feature extraction capabilities of graph neural networks for complex interaction relationships. It has broken through the limitations of high cost and difficult coverage of space in traditional simulation calculations, made up for the deficiencies of single-process machine learning models, improved the integrity and work efficiency of defect prediction, and provided a more accurate tool for the yield optimization of advanced processes [7]. The key technology is multi-scale feature fusion, combined with transistor-level geometric parameters and wafer-level process parameters. For instance, in the prediction of contact hole size deviation defects, the prediction accuracy is 32% higher than that of traditional physical simulation, and it can identify the failure of too small hole diameter caused by "lithography miniaturization + anisotropic superposition of etching" that was overlooked by traditional methods [8].

2.2. Other key technologies for AI-enabled chip verification

In the optimization of lithography processes, reinforcement learning, by combining dynamic decision-making with process window constraints, has developed a set of parameter adaptive adjustment strategies to address the coupling problem of thermal effects and ESD protection. This strategy can achieve efficient global optimization in a process environment with multiple variables, strong constraints, and nonlinear correlations. For instance, after applying this method in TSMC's N7+ process, the lithography yields significantly increased by 18% [9]. However, this method also faces challenges such as the reliance on experience in the design of the reward function, the high cost of parameter adjustment in the early stage, and the limited generalization ability across processes. In the field of test generation and aging prediction, random forests, with their multi-source feature screening and defect classification capabilities, provide an effective path for test vector optimization. Intel has automatically identified key features from multi-source data through feature importance assessment and nonlinear combinatorial analysis, replacing the previous test generation method that relied on experience and successfully reducing the number of test vectors by 42% [10]. This method can balance test efficiency and coverage without complex engineering, and can handle redundant information in multi-source data. However, it is relatively sensitive to extreme outliers, and its

accuracy may be limited under high-dimensional complex features. Moreover, the model's interpretability is relatively weak. In addition, deep neural networks have brought about significant breakthroughs in the acceleration of physical field simulation, especially in IR voltage drop analysis and electromigration prediction. The DSO.ai tool developed by Synopsys fuses prior knowledge of physical fields through end-to-end mapping to build a simulation proxy model, thus skipping the traditional numerical iteration process and significantly reducing the simulation time of 3nm chips from 72 hours to 18 hours [11]. This data-driven approach effectively breaks through the time bottleneck of nanoscale multi-field coupled simulation, but its performance largely depends on the coverage of training data. Errors may occur in some scenarios, and the model needs to be continuously updated, with relatively limited interpretability.

3. Discussion and Analysis

Table 1 presents a comparative analysis of three technologies, namely GNN, RL and DNN accelerated simulation, in four dimensions: data requirements, real-time performance, interpretability and applicable scenarios.

Table 1. Comparison of different dimensions

Dimension	GNN	RL	DNN accelerated simulation
Data requirements	High (Process parameters need to be marked)	Medium (Reward function design required)	Low (can directly process raw data)
Real-time performance	Medium (minute-level reasoning)	Low (requires multiple iterations)	High (real-time simulation)
Interpretability	High (Path visualization)	Low (Black box decision-making	(Feature Importance Analysis)
Applicable scenarios	Microscopic defect prediction	Macroscopic parameter optimization	Physical field acceleration simulation

The robustness issue of AI verification, as measured by Samsung Foundry, shows that when the fluctuation of process parameters exceeds ±7σ, the failure probability of existing AI models surges [12]. However, the solution proposed by MIT is to construct an adversarial training set containing process anomaly data, which enhances the model's robustness by 41% in 3nm process validation [13]. The latest research by the MIT team shows that the DNN model has the defect of overfitting the heat sink layout and ignoring the power network topology when predicting IR voltage drop [14]. However, industry representatives pointed out that explainability should not come at the expense of efficiency. TSMC's N3 process strikes a balance between accuracy and explainability by introducing an explainability threshold [15].

Under the deep integration of artificial intelligence and chip verification, workers in the microelectronics field need to build a three-in-one composite capability system of algorithms, processes, and designs. To break through the traditional toolchain thinking, master core technologies such as graph neural network modeling and reinforcement learning, for instance, solve the problem of lithography error propagation prediction through the multi-scale feature fusion method proposed by MIT, and establish a cross-scale cognitive framework that can deeply understand the microscopic characterization mechanism of device-level process deviations. The simulation efficiency of 3nm chips can also be improved through tools such as Synopsys DSO.ai.What is important is that, in the rapid development of technology, it is necessary to cultivate the awareness of technical ethics simultaneously. It is essential to enhance the robustness of the model under extreme process fluctuations of ±7σ through adversarial training, and also to balance the accuracy and interpretability indicators in model optimization, such as referring to the experience of setting interpretability

thresholds in TSMC's N3 process. This multi-dimensional expansion of technical capabilities will drive practitioners to transform from single tool users to the role of AI and chip hybrid verification system architecture, becoming the core force promoting the intelligence of EDA and the development of the open-source ecosystem.

4. Conclusion

This study, through a systematic review of the application status and development trends of artificial intelligence technology in the field of chip verification, combined with the core course knowledge system of microelectronics and industrial cases, has demonstrated that AI technology has played a significant role in enhancing the efficiency of chip verification, reducing the rate of missed detection of process defects, and accelerating the simulation of physical effects. To address the challenges of process deviation sensitivity and data fragmentation in advanced manufacturing processes, graph neural networks have achieved efficient prediction of microscopic defects such as lithography errors and doping fluctuations through multi-scale feature fusion and dynamic graph update mechanisms. Reinforcement learning and deep neural networks have respectively demonstrated strong engineering adaptability in the optimization of lithography parameters and the acceleration of physical field simulation. Research has confirmed that AI technology can increase the accuracy rate of process-related defect prediction to over 85% and boost the simulation efficiency of 3nm processes to more than four times that of traditional methods. Although current AI models face robustness bottlenecks in extreme process fluctuation scenarios, a dynamic balance between accuracy and efficiency has been initially achieved through the design of adversarial training and interpretability constraint mechanisms. Future work can further explore the collaborative mechanism of multimodal AI models in cross-scale verification tasks and promote the practical implementation of AI-driven verification processes in open-source chip projects.

References

[1] Li X, Zhang Y, Liu C, et al. Cross-layer defect propagation prediction with transformer-based self-attention. In: Proceedings of the 44th International Conference on Software Engineering (ICSE). 2022: 1125–1136.

[2] Li W, Curtis S, Wang Y, et al. PhysX-AI: Accelerating electromigration prediction with neural-network-enhanced finite element analysis. In: Proceedings of the 49th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH). 2023: 1–12.

[3] Kim J, Lee S, Park H, et al. AutoV: Federated learning for cross-domain data fusion in semiconductor verification. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–10.

[4] Wang Y, Chen T, Li Z, et al. AI-driven verification toolkit: Integrating PyTorch-based process parameter optimization for RISC-V rapid validation. In: Proceedings of the 61st Design Automation Conference (DAC). 2024: 1–10.

[5] Zhou J, Zhang L, Chen K, et al. PGNet: Process variation-aware transistor performance prediction with graph neural networks. In: Proceedings of the 59th Design Automation Conference (DAC). 2022: 1–10.

[6] Chen Y, Lee K, Wang L, et al. RL-Litho: Reinforcement learning-driven optimization of lithography process parameters for sub-3nm node yield enhancement. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–12.

[7] Singh A, Patel R, Lee S, et al. CoupledProc-GNN: Unveiling lithography-etch defect synergies via graph neural networks for advanced node manufacturing. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–11.

[8] Xie Z, Ma Y, Yu B, et al. Graph neural networks for coupled lithography-etch defect prediction. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2021, 40(8): 1681–1694.

[9] Huang C, Lee J, Chen W, et al. Industrial deployment of coupled AI-process modules for lithography yield enhancement in TSMC N7+ node. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–12.

[10] Shen Y, Zhang L, Lee J, et al. RF-driven test vector compression for aging-aware validation in Intel 10nm processors. In: Proceedings of the 60th Design Automation Conference (DAC). 2022: 1–10.

[11] Chen G, Lee H, Wang Q, et al. DNN-accelerated electrothermal simulation for 3nm IR drop and electromigration analysis with Synopsys DSO.ai. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–12.

[12] Kim S, Park J, Lee H, et al. Robustness limits of AI-driven process control models under extreme σ fluctuations: A Samsung Foundry 3nm validation study. In: Proceedings of the 61st Design Automation Conference (DAC). 2024: 1–10.

[13] Park J, Kim S, Lee H, et al. Adversarial process anomaly training for robust 3nm process validation: Enhancing model resilience against extreme σ fluctuations. In: Proceedings of the 61st Design Automation Conference (DAC). 2024: 1–11.

[14] Li X, Wang Q, Zhang Y, et al. On the feature overfitting of DNN models for IR drop prediction: The neglect of power network topology defects. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–10.

[15] Hsieh C, Chen T, Lee S, et al. Explainable thresholding for accuracy-interpretability balance in TSMC N3 process control models. In: Proceedings of the 60th Design Automation Conference (DAC). 2023: 1–10.