Efficient Modulated State Space Model for Mixed-Type Wafer Defect Pattern Recognition

Mu Nie
School of Integrated Circuits
Anhui Polytechnic University
Wuhu, China
niemu@ahpu.edu.cn

Shidong Zhu
School of Integrated Circuits
Anhui Polytechnic University
Wuhu, China
2981937589@qq.com

Aibin Yan
School of Microelectronics
Hefei University of Technology
Hefei, China
abyan@mail.ustc.edu.cn

Cheng Zhuo
College of Integrated Circuits
Zhejing University
Hangzhou, China
czhuo@zju.edu.cn

Xiaoqing Wen
Department of Computer Science and Networks
Kyushu Institute of Technology
Fukuoka, Japan
wen@cse.kyutech.ac.jp

Tianming Ni
School of Integrated Circuits
Anhui Polytechnic University
Wuhu, China
timmyni126@126.com

***Abstract*—Accurate and efficient wafer defect detection is crucial in semiconductor manufacturing to maintain product quality and optimize yield. Traditional methods struggle with the complexity and diversity of modern wafer defect patterns. While deep learning approaches are effective, they are often resource-intensive, posing challenges for real-time deployment in industrial settings. To solve these problems, we propose an Efficient Modulated State Space Model (EM-SSM) for mixed-type wafer defect recognition, optimized with knowledge distillation to balance accuracy and efficiency. Our framework captures size-dependent relationships and improves defect-specific feature representation to recognize complex defects precisely. Specifically, we introduce an efficient directional modulation mechanism to refine spatial recognition of defect patterns. To further improve inference efficiency, we propose a deep-to-shallow distillation method that transfers knowledge from deeper networks to lighter networks, reducing inference time without compromising classification accuracy. Experimental results on the MixedWM38 wafer dataset with 38 defect types show that our model achieves 99.0% accuracy, outperforming traditional methods in both accuracy and efficiency. Our model offers a scalable solution for modern semiconductor defect detection.**

***Index Terms*—Mixed-type Wafer Map, State Space Model, Defect Pattern, Failure Mode Analysis**

I. Introduction

Wafer manufacturing is a critical phase in semiconductor production, involving numerous intricate steps. After production, electrical testing is performed on each chip using the probe card, and the results are documented in wafer maps. These maps are essential for identifying defects, which are often linked to specific manufacturing processes. Early defect detection is vital for diagnosing process issues, improving yield, and preventing large-scale production failures [1]. For instance, central defects may result from uniformity variations during chemical-mechanical planarization (CMP), while scratch defects can be caused by particle accumulation and pad wear. Edge-ring defects are frequently associated with misalignment in the storage node (ND) process [2]. Consequently, accurate and efficient wafer defect pattern detection is critical for optimizing yield and ensuring production reliability.

The rapid expansion of integrated circuits and semiconductor manufacturing have led to an exponential increase in data generation. This surge in data has driven significant advancements in defect recognition algorithms, particularly those based on deep learning techniques [1], [3]. Convolutional Neural Networks (CNNs), known for extracting high-level features from complex datasets automatically, have garnered widespread attention in wafer defect recognition. CNNs enable end-to-end training, significantly improving classification accuracy and facilitating efficient and precise identification of various defects in wafer maps [4], [5]. Moreover, the integration of Electronic Design Automation (EDA) tools into these workflows has further enhanced automation, enabling faster and more accurate design iterations while minimizing manual intervention in defect analysis and correction. This integration has streamlined the overall process, making defect detection more efficient and reliable [6].

As the density of integrated circuits continues to increase, defects on wafers frequently combine to form more complex, mixed-type patterns. These patterns are significantly more challenging to detect compared to simpler ones, posing substantial difficulties for existing defect recognition methods [7], [8]. Despite the critical importance of accurately detecting these patterns, few studies have effectively applied CNNs or transformer architectures to address this challenge [4], [8]. Several limitations persist in previous approaches: First, CNN-based feature extraction methods are highly sensitive to noise and often struggle to capture the global structure of wafer maps, resulting in suboptimal performance in detecting mixed-type defects [8]. Second, although vision transformers (ViTs) have been introduced [4], [9], their reliance on CNN-based blocks and the need for large datasets lead to increased computational

This work was supported in part by the National Natural Science Foundation of China under grant (No. 62174001, 62311540021). The corresponding author of this paper is Tianming Ni (Email: timmyni126@126.com).

978-3-9826741-0-0/DATE25/© 2025 EDAA

overhead. PaLM [10] leverages point clouds in two-dimensional space to extract complex geometric features, but it requires extensive pre-training and fine-tuning to achieve optimal performance. This dependence on large-scale pre-trained models presents additional challenges in applying PaLM to wafer defect detection, particularly without significant computational resources. Moreover, techniques such as DeiT [11], which employ a teacher-student distillation strategy, demonstrate promise in knowledge transfer but face practical challenges when applied to wafer defect detection.

In this paper, to address the challenges in detecting complex, mixed-type wafer defects, we present an Efficient Modulated State Space Model, optimized through knowledge distillation for enhanced computational efficiency and accuracy. The sophisticated model dynamically captures size-dependent relationships, improving defect-specific feature representation, while the directional modulation mechanism adaptively adjusts fusion weights across anisotropic scanning directions, enhancing spatial understanding of defect patterns. By transferring knowledge from deeper networks to a more efficient architecture, our approach reduces inference time without compromising classification accuracy.

The main contributions of this paper are as follows:

- We propose an adaptive direction-modulated state space model that effectively captures size-dependent relationships, improving the identification of mixed-type wafer defects.
- We develop a deep-to-shallow distillation method to transfer knowledge from deep networks into a lightweight model, reducing inference time without sacrificing accuracy.
- Our method achieved leading performance on the MixedWM38 dataset, balancing both accuracy and efficiency, accurately recognizing 38 defect types and capturing intricate defect patterns.

The rest remainder of this paper is organized as follows: Section II introduces the preliminaries on mixed-type wafer defect pattern recognition. Section III describes the proposed framework, including the modulated state space model and knowledge distillation process. Section IV presents the experimental setup and results. Finally, conclusions are drawn in Section V.

II. Preliminaries

We use the MixedWM38 wafer map dataset [3], comprising 38,015 wafer maps with 38 distinct single and mixed-type defect patterns. Each 52×52 wafer map represents electrical test results for individual dies, with three states: No die is present at this location (marked as 0). The die is present and passes the electrical test (marked as 1). The die is present but fails the test (marked as 2).

Figure 1 shows examples of both single-type and mixed-type wafer defect patterns. The dataset includes eight single-type defects (e.g., *Center*, *Donut*, *Near_Full*) and a variety of mixed-type defects (including 13 two-mixed-type defects, 12 three-mixed-type defects, and 4 four-mixed-type defects). The

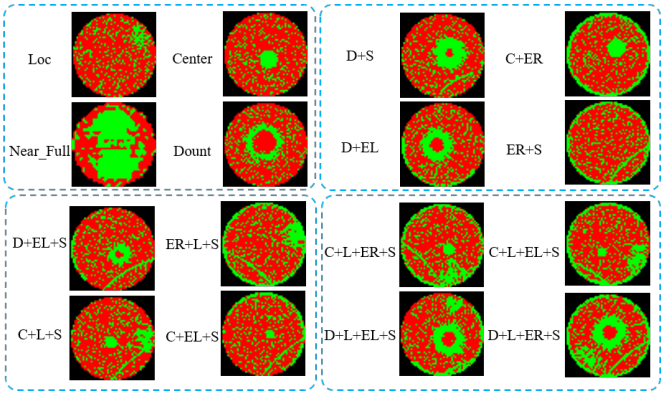

Fig. 1. Single and mixed-type wafer patterns in the benchmark. To better highlight the defects, we applied color processing to the data.

complexity of mixed-type defects stems from their combination of distinct failure modes within a single wafer, often the result of multi-stage process variations in modern semiconductor fabrication, which involves hundreds of process steps.

The detection task involves identifying defective dies and recognizing defect patterns that provide insight into the manufacturing process. Our efficient modulated state space model dynamically captures size-dependent relationships and adapts to complex spatial distributions using direction modulation, improving computational efficiency via knowledge distillation.

III. Methodology

A. Overview of Efficient Modulated State Space Model

The proposed Efficient Modulated State Space Model (EMSSM) is tailored for wafer defect recognition using a Direction Modulated State Space Model (DMSS). The overall architecture is shown in Figure 2. The input wafer map $I \in \mathbb{R}^{H\times W\times 3}$ is divided into patches, generating a feature map. A series of DMSS blocks are applied at progressively lower resolutions, using down-sampling to capture spatial patterns efficiently.

Each DMSS block consists of two main components: a 2D-Selective-Scan (SS2D) module and a Direction Modulated (DM) structure, designed for computational efficiency. The DMSS block, adapted from the Selective Scan Mechanism architecture [12], allows the model to capture spatial dependencies in wafer defect images.

B. Formulation of State Space Models

State Space Models (SSMs) originate from Kalman filters and are a type of linear time-invariant (LTI) system that maps input signals $u(t) \in \mathbb{R}$ to output responses $y(t) \in \mathbb{R}$ through hidden states $h(t) \in \mathbb{R}^N$. The continuous-time dynamics of an SSM can be represented by the following linear ordinary differential equations (ODEs):

$$h'(t) = Ah(t) + Bu(t), \tag{1}$$

$$y(t) = Ch(t) + Du(t), \tag{2}$$

where $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$, $C \in \mathbb{R}^{1\times N}$, and $D \in \mathbb{R}^1$ are system matrices that transform input signals into output responses through the hidden state.

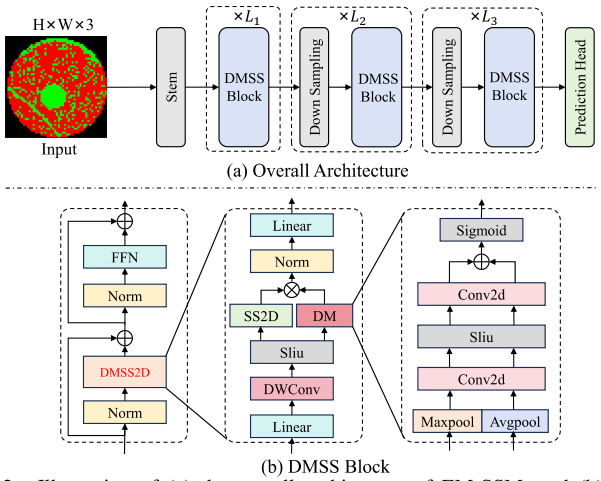

Fig. 2. Illustration of (a) the overall architecture of EM-SSM, and (b) the structure of the DMSS block.

To integrate continuous-time SSMs into neural architectures, discretization is necessary. Given a time interval $[t_a, t_b]$, the hidden state $h(t_b)$ can be derived by solving the ODE:

$$h(t_b) = e^{A(t_b - t_a)} h(t_a) + \int_{t_a}^{t_b} e^{A(t_b - \tau)} B u(\tau)\, d\tau. \quad (3)$$

Using a time step Δ, the discrete form is approximated as:

$$h_b = e^{A\Delta} h_a + \sum_{i=a}^{b-1} e^{A\Delta} B u_i. \quad (4)$$

This approximation employs the Zero-Order Hold (ZOH) method, which is widely used for SSM-based models [12].

C. Selective Scan Mechanism

To achieve the spatial relation of the wafer, we adopt the 2D-Selective-Scan (SS2D) mechanism [12], depicted in Figure 3. Unlike 1D-based SSMs optimized for temporal sequences, SS2D is engineered to manage the structured spatial patterns intrinsic to visual data, whereas conventional SSMs tend to compromise spatial information due to their inherently sequential operation. The SS2D architecture is uniquely adapted for visual data, incorporating a dynamic, input-dependent selection method that modifies weighting parameters in real time based on the characteristics of the input. This adaptation enables precise handling of both local textures and overarching structural elements essential in applications like wafer defect detection.

The operation of SS2D encompasses a three-stage process: cross-scan, selective scanning through input-dependent selection mechanism blocks, and cross-merge. Initially, the model unfolds input data into sequences along multiple traversal paths (Cross-Scan), each processed in parallel by its respective input-dependent selection mechanism block. These blocks are crucial for the selective scanning stage, dynamically adjusting weights to effectively manage complex patterns across varied regions. Following processing, the sequences are restructured and amalgamated (Cross-Merge), generating an output map that preserves essential spatial dependencies. This method effectively improves the recognition of mixed-type wafer defects while maintaining computational efficiency through the use of linear-complexity scanning algorithms, balancing performance and resource utilization.

D. Direction Modulation

Although [12] employs a 4-directional scanning mechanism, it assigns equal importance to each direction, meaning the weights of cross-merge are uniform. This may result in insufficient perception capabilities for defects with specific spatial distributions. To mitigate this problem, we propose the Direction Modulation (DM) operator, which re-calibrates direction-wise feature responses to enhance network representation power. Specifically, we first apply global average pooling to compute channel statistics as shown in (5).

$$\mathrm{MA}(\mathbf{X}) = \mathrm{Concat}(\mathrm{MaxPool}(\mathbf{X}), \mathrm{AvgPool}(\mathbf{X})), \quad (5)$$

where $\mathbf{X}$ is the input tensor, MaxPool represents the global max pooling operation, and AvgPool represents the global average pooling operation. These operations are concatenated to form the channel statistics. The affinity calculation is performed as shown in (6).

$$\mathbf{X}_{\mathrm{DM}} = \mathrm{DM}(\mathbf{X}) = \sigma(Conv_2 \cdot \delta(Conv_1 \cdot \mathrm{MA}(\mathbf{X}))). \quad (6)$$

Here, δ and σ denote non-linearity functions, and $Conv_1$ and $Conv_2$ are convolution layers. The $\mathbf{X}_{\mathrm{DM}}$ score assigns importance to each channel, facilitating the recalibration of the output of the DM operator.

E. Deep-to-Shallow Distillation Framework

In this work, we employ a deep-to-shallow distillation strategy, transferring knowledge from a larger, more complex teacher model to a streamlined student model (as illustrated in Figure 4). The key difference between the teacher and student models lies in their structure, where the teacher utilizes a three-stage architecture with [2,2,5], while the student model adopts a more lightweight three-stage setup with [1,1,1].

1) Prediction Alignment.: In general, deep learning models rely on ground truth labels to constrain the outputs during the training phase, to have the model's output as similar as possible to the ground truth. In this work, we use the commonly used multi-class cross-entropy loss to calculate the loss:

$$L_{\mathrm{CE}} = -\sum_{i=1}^{C} y_i \log(p_i), \quad (7)$$

where C represents the number of classes, which is 38 in this case, and y_i and p_i represent the ground truth and predicted output for the i class, respectively.

However, previous studies [11], [13] have shown that the soft labels provided by the teacher network are more beneficial for the gradient backpropagation learning of the student network compared to the one-hot ground truth labels. Therefore, we use the output of the teacher network for prediction alignment. We utilize Kullback-Leibler (KL) divergence to capture the difference between the soft probability distributions of the two models:

$$L_{\mathrm{KL}} = KL(P_s \| P_t). \quad (8)$$

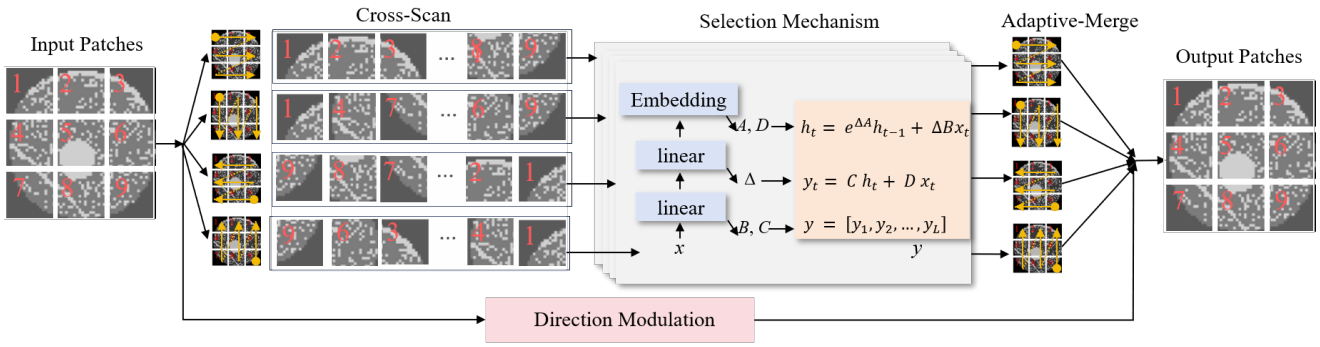

Fig. 3. Illustration of the Direction Modulated 2D-Selective-Scan (DMSS2D) mechanism. Input patches are navigated along four distinct scanning paths (Cross-Scan), with each pathway processed independently by dedicated input-dependent selection mechanism blocks. Finally, the four-directional scanning results are adaptively fused based on the spatial characteristics of the input features.

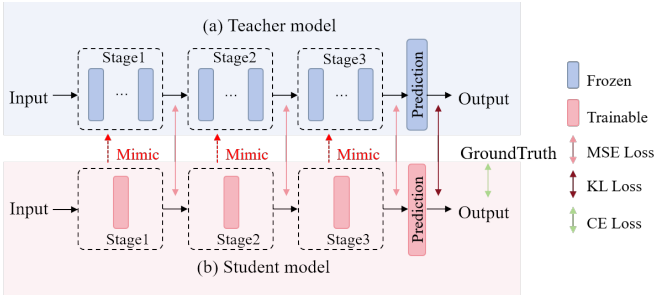

Fig. 4. Teacher-student model architecture used in the distillation process.

This loss ensures that the student model not only mimics the teacher's final decision-making process but also captures the nuances in the soft-label distribution provided by the teacher.

2) Feature Mimic: The feature mimic encourages the student model to learn from the intermediate feature maps of the teacher model. Using a mapping function $\Psi(\cdot)$, we compress feature maps into unified response maps, and the student learns to approximate the teacher's feature responses using mean squared error (MSE):

$$L_{\text{MSE}} = \sum \left(\Psi(F_t) - \Psi(F_s)\right)^2. \tag{9}$$

Here, it $\Psi(\cdot)$ denotes the function that compresses the teacher model's feature map F_t and the student model's feature map F_s into single-channel response maps.

3) Total Distillation: The total distillation loss combines the three components to achieve deep-to-shallow knowledge transfer:

$$L_{\text{total}} = \lambda_1 L_{\text{CE}} + \lambda_2 L_{\text{KL}} + \lambda_3 L_{\text{MSE}}, \tag{10}$$

where λ_1, λ_2, and λ_3 are hyperparameters that adjust the balance between the different loss components, allowing for fine-tuning based on specific requirements.

IV. EXPERIMENTAL RESULTS

A. Experimental Settings

Our framework was implemented using PyTorch and evaluated on an Nvidia RTX 3090 GPU platform. The MixedWM38 benchmark dataset consists of 38,015 wafer images from 46,293 batches, introduced in Section II, all annotated by fab experts. Each pattern category contains an equal number of samples, except for *Donut*, *C+L+EL+S*, and *D+L+EL+S*, which have slightly different counts. The wafer maps were split into 70% for training and 30% for testing.

For training details, our models were trained for 120 epochs using the Adam optimizer with an initial learning rate of 0.001, which was decayed by a factor of 0.1 every 40 epochs. We used a batch size of 128 to optimize training efficiency. Knowledge distillation was applied with the following loss weights: the KL-divergence weight of 0.05, the MSE weight of 0.01, and the ground truth target weight of 1.0.

B. Comparisons Against the Latest Models

In comparison with the latest models from MSF-Trans [4], SAM [14] , and PaLM [10], our proposed model consistently achieves superior precision across various defect patterns, underscoring its advancements in handling complex and mixed-type defects. As shown in Table I, for simpler defect patterns, such as *Center* and *Donut*, our model achieves slight but notable precision improvement, reaching 0.995 and 0.988, respectively, surpassing PaLM's 0.974 and 0.984. In more complex mixed-defect scenarios like *C+L* and *D+ER+S*, our model achieves superior precisions of 0.990 and 1.000, outperforming PaLM's 0.987 and 0.967. These results demonstrate the efficacy of our directional modulation and adaptive feature fusion techniques in accurately identifying intricate defect patterns, particularly in challenging detection tasks.

In challenging scenarios such as *D+L+ER+S*, our model achieves a perfect precision of 1.000, substantially outperforming PaLM's 0.973. This demonstrates the model's robustness in handling intricate defect combinations, driven by advanced techniques like transfer learning and directional modulation. These results underscore the consistent performance gains of our approach, making it well-suited for practical deployment in wafer defect detection, where high precision is essential for optimizing yield and reducing costs.

C. Ablation Studies

1) Direction Modulation: The ablation study demonstrates the benefits of direction modulation in enhancing both computational efficiency and accuracy. Table II shows that our models achieve higher accuracy and significantly improve computational efficiency compared to the PaLM baseline. The

TABLE I
COMPARISON OF PRECISION, RECALL, AND F1 SCORE FOR DIFFERENT MODELS AND PATTERNS.

Pattern	MSF-Trans (TSM'22 [4])			SAM (ICCV'23 [14])			PaLM(DATE'24 [10])			Our Model		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Normal	0.974	1.000	0.987	0.974	0.993	0.983	0.993	1.000	0.997	1.000	0.985	0.992
Center	0.936	0.973	0.954	0.905	0.957	0.930	0.974	1.000	0.987	0.995	0.985	0.990
Donut	0.967	0.973	0.970	0.976	0.947	0.961	0.984	0.997	0.990	0.988	0.990	0.989
Edge_Loc	0.976	0.937	0.956	0.983	0.937	0.959	0.990	0.997	0.993	0.985	0.995	0.990
Edge_Ring	0.938	0.963	0.951	0.930	0.970	0.949	0.987	0.990	0.988	0.980	0.995	0.988
Loc	0.961	0.977	0.969	0.970	0.957	0.963	0.983	0.993	0.988	0.976	1.000	0.988
Near_Full	0.992	0.950	0.971	0.961	0.950	0.956	0.996	0.996	0.996	0.985	0.985	0.985
Scratch	0.954	0.970	0.962	0.948	0.970	0.959	0.997	0.997	0.997	1.000	0.965	0.982
Random	0.759	0.911	0.828	0.729	0.778	0.753	0.977	0.933	0.955	0.990	0.980	0.985
C+EL	0.954	0.960	0.957	0.950	0.947	0.948	0.987	0.987	0.987	0.990	0.990	0.990
C+ER	0.951	0.977	0.964	0.918	0.977	0.947	0.980	0.980	0.980	0.985	0.990	0.988
C+L	0.960	0.963	0.962	0.948	0.963	0.955	0.987	0.983	0.985	0.990	1.000	0.995
C+S	0.930	0.973	0.951	0.910	0.973	0.940	0.980	0.990	0.985	0.995	0.985	0.990
D+EL	0.944	0.957	0.950	0.938	0.957	0.947	0.990	0.980	0.985	0.995	0.995	0.995
D+ER	0.944	0.957	0.950	0.938	0.957	0.947	0.990	0.980	0.985	0.990	0.995	0.993
D+L	0.935	0.953	0.944	0.934	0.947	0.940	0.977	0.980	0.978	0.995	0.995	0.995
D+S	0.956	0.950	0.953	0.963	0.960	0.962	0.993	0.990	0.992	0.995	0.995	0.995
EL+L	0.912	0.970	0.940	0.877	0.950	0.912	0.967	0.983	0.975	1.000	1.000	1.000
EL+S	0.972	0.933	0.952	0.946	0.933	0.940	0.997	0.980	0.988	0.985	0.960	0.970
ER+L	0.927	0.967	0.946	0.922	0.930	0.926	0.964	0.993	0.979	1.000	0.985	0.992
ER+S	0.960	0.960	0.960	0.959	0.930	0.944	0.983	0.987	0.985	1.000	1.000	1.000
L+S	0.963	0.963	0.963	0.964	0.970	0.967	0.984	0.997	0.990	0.970	0.980	0.975
C+EL+L	0.979	0.953	0.966	0.976	0.943	0.959	0.997	0.970	0.983	0.980	1.000	0.990
C+EL+S	0.988	0.977	0.982	0.986	0.965	0.976	0.997	0.990	0.993	0.980	0.995	0.988
C+ER+L	0.873	0.940	0.905	0.869	0.930	0.899	0.951	0.980	0.966	0.990	0.995	0.993
C+ER+S	0.973	0.977	0.975	0.969	0.943	0.956	0.983	0.993	0.988	0.985	0.965	0.975
C+L+S	0.983	0.937	0.959	0.983	0.940	0.961	0.986	0.970	0.978	0.995	0.995	0.995
D+EL+L	0.937	0.937	0.937	0.927	0.927	0.927	0.990	0.977	0.983	0.995	0.975	0.985
D+EL+S	0.951	0.903	0.926	0.944	0.903	0.923	0.973	0.973	0.973	0.961	0.990	0.975
D+ER+L	0.970	0.957	0.963	0.962	0.933	0.948	0.993	0.987	0.990	1.000	0.990	0.995
D+ER+S	0.884	0.940	0.911	0.877	0.923	0.899	0.967	0.970	0.968	1.000	0.995	0.998
D+L+S	0.968	0.917	0.942	0.962	0.917	0.939	0.980	0.973	0.977	0.985	1.000	0.993
EL+L+S	0.978	0.910	0.943	0.961	0.893	0.926	0.997	0.963	0.980	0.990	0.975	0.982
ER+L+S	0.935	0.917	0.926	0.933	0.923	0.928	0.980	0.960	0.970	0.990	0.995	0.993
C+L+EL+S	0.959	0.930	0.944	0.941	0.910	0.925	0.980	0.967	0.973	0.994	0.994	0.994
C+L+ER+S	0.932	0.863	0.896	0.926	0.870	0.897	0.963	0.950	0.956	0.995	1.000	0.998
D+L+EL+S	0.965	0.910	0.937	0.947	0.887	0.916	0.980	0.970	0.975	1.000	0.966	0.983
D+L+ER+S	0.944	0.953	0.949	0.926	0.953	0.939	0.971	0.990	0.980	0.985	1.000	0.993
Average	0.946	0.949	0.948	0.963	0.967	0.965	0.982	0.982	0.982	0.990	0.990	0.989

TABLE II
COMPARISON OF MODEL PARAMETERS AND PERFORMANCE

Model	FLOPs (G)	Params (M)	FPS (img/s)	Test Acc (%)
PaLM([10])	-	-	110.99	98.2
Student-DM	0.1067	3.45	344.09	98.3
Teacher+DM	0.1906	9.11	116.75	99.0
Student+DM	0.1088	5.00	279.03	99.0

student without DM model is particularly efficient, with fewer parameters and faster processing speeds, while maintaining strong accuracy. Overall, the student with DM model shows that our approach successfully balances performance and efficiency, making it suitable for real-time wafer defect detection.

2) Knowledge Distillation: To evaluate the effectiveness of Knowledge Distillation (KD) in enhancing model performance, we conducted an ablation study comparing three different models: the teacher model, a student model trained without KD, and a student model trained with KD in table5. The application of KD leads to a noticeable enhancement in all evaluated metrics. For instance, precision increased by 0.008, recall by 0.008, and F1-score by 0.007 when KD was applied. This suggests that KD effectively transfers valuable knowledge from the teacher model, allowing the student model to perform better even with a reduced architecture. The observed improvements highlight KD's potential to optimize model performance by leveraging the expertise of more complex models.

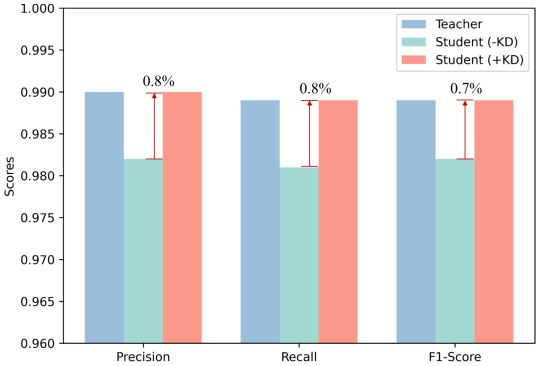

Fig. 5. The effectiveness of Knowledge Distillation.

D. Visualization

The visualization results in Figure 6 illustrate the improved defect detection performance of our model with the integration of the Direction Modulation (DM) module. For both simple defect types (e.g., *Center* and *Donut*) and more complex mixed-type defects, the activation maps (shown on the right) highlight the model's enhanced focus on critical defect regions, as indicated by the red arrows. The DM module refines the model's attention, leading to more precise and concentrated activations, particularly in scenarios involving complex mixed defects.

The refined activation maps show the model's improved capacity to identify multiple defect regions simultaneously. In the lower rows, for instance, the model with DM accurately detects overlapping or interacting defects, as highlighted by the red arrows. This demonstrates the module's robustness in handling intricate defect patterns. Through deeper layers, the model progressively refines its detection capability, concentrating on key areas that contribute to higher detection accuracy, particularly in challenging defect patterns.

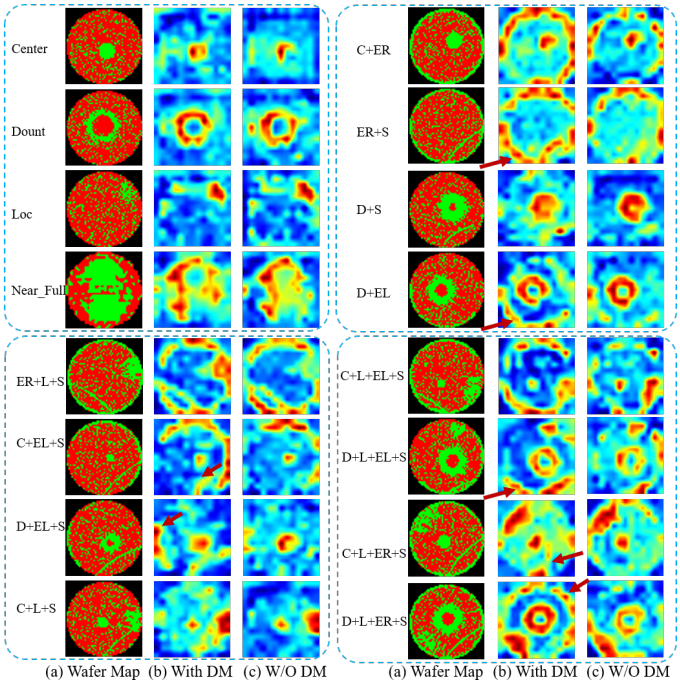

Fig. 6. Visualization of defect detection activations with the DM module, showing the enhanced focus on critical defect areas in both single and mixed defect scenarios.

V. CONCLUSIONS

This paper has presented an Efficient Modulated State Space Model (EM-SSM) for wafer defect pattern recognition, which demonstrates superior performance in both single-type and mixed-type defect detection tasks. By integrating direction modulation and adaptive feature fusion, our model significantly improves defect localization and classification, particularly in complex scenarios. Comprehensive experiments indicate that EM-SSM outperforms state-of-the-art methods like PaLM in terms of accuracy, precision, and computational efficiency. The incorporation of the direction modulation module proves especially effective in enhancing feature extraction for mixed defects. These results suggest that EM-SSM not only achieves state-of-the-art performance on standard benchmarks but also shows strong potential for deployment in real-world semiconductor manufacturing. Future work will explore the extension of this framework to broader defect categories and its integration into real-time industrial inspection systems.

REFERENCES

[1] H. Geng, Q. Sun, T. Chen, Q. Xu, T.-Y. Ho, and B. Yu, "Mixed-type wafer failure pattern recognition (invited paper)," in *2023 28th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2023, pp. 727–732.

[2] J. Yan, Y. Sheng, and M. Piao, "Semantic segmentation-based wafer map mixed-type defect pattern recognition," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 11, pp. 4065–4074, 2023.

[3] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 4, pp. 587–596, 2020.

[4] Y. Wei and H. Wang, "Mixed-type wafer defect recognition with multi-scale information fusion transformer," *IEEE Transactions on Semiconductor Manufacturing*, vol. 35, no. 2, pp. 341–352, 2022.

[5] S. Zhao, Z. Zhu, X. Li, and Y.-C. Chen, "Robust wafer classification with imperfectly labeled data based on self-boosting co-teaching," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 7, pp. 2214–2226, 2023.

[6] K. C.-C. Cheng, K. Shu-Min Li, A. Y.-A. Huang, J.-W. Li, L. L.-Y. Chen, N. Cheng-Yen Tsai, S.-J. Wang, C.-S. Lee, L. Chou, P. Y.-Y. Liao, H.-C. Liang, and J.-E. Chen, "Wafer-level test path pattern recognition and test characteristics for test-induced defect diagnosis," in *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020, pp. 1710–1711.

[7] M. Nie, W. Jiang, W. Yang, S. Wang, X. Wen, and T. Ni, "Enhancing defect diagnosis and localization in wafer map testing through weakly supervised learning," in *2023 IEEE 32nd Asian Test Symposium (ATS)*, 2023, pp. 1–6.

[8] X. Zhang, Z. Jiang, H. Yang, Y. Mo, L. Zhou, Y. Zhang, J. Li, and S. Wei, "Dmwmnet: A novel dual-branch multi-level convolutional network for high-performance mixed-type wafer map defect detection in semiconductor manufacturing," *Computers in Industry*, vol. 161, p. 104136, 2024.

[9] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9992–10 002.

[10] H. He, G. Kuang, Q. Sun, and H. Geng, "Palm: Point cloud and large pre-trained model catch mixed-type wafer defect pattern recognition," in *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2024, pp. 1–6.

[11] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training data-efficient image transformers & distillation through attention," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 10 347–10 357.

[12] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "Vmamba: Visual state space model," *arXiv preprint arXiv:2401.10166*, 2024.

[13] G. Hinton, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[14] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 3992–4003.