

SDM-PEB: Spatial-Depthwise Mamba for Enhanced Post-Exposure Bake Simulation

Ziyang Yu¹, Peng Xu¹, Zixiao Wang¹, Binwu Zhu², Qipan Wang³, Yibo Lin³, Runsheng Wang³, Bei Yu¹, Martin Wong⁴
¹The Chinese University of Hong Kong ²Southeast University&NCTIEDA ³Peking University ⁴Hong Kong Baptist University

Abstract—The post-exposure bake (PEB) process is a critical step in semiconductor lithography, directly impacting resist profile accuracy and circuit pattern fidelity. Precise modeling of PEB is essential for controlling photoacid diffusion and inhibitor reactions. In this paper, we introduce SDM-PEB, an advanced modeling framework designed to enhance the accuracy of PEB simulations by capturing both intra-layer spatial dependencies and inter-layer depthwise interactions. Leveraging a unique hierarchical feature extractor with overlapped patch merging and efficient self-attention, our approach effectively captures both coarse and fine features at multiple scales. The spatial-depthwise Mamba-based attention unit, centered on a customized selective scan and structured state space model, efficiently captures spatial and depthwise dependencies, enabling precise 3D PEB simulation. Additionally, a PEB focal loss and differential depth divergence regularization term improve the sensitivity to both spatial and depthwise variations, addressing inherent data imbalances in 3D PEB simulations. Our framework is validated with commercial rigorous model, and experimental results demonstrate that the SDM-PEB outperforms previous methods in accuracy and efficiency.

I. INTRODUCTION

Lithography simulation plays a vital role in semiconductor manufacturing, providing precise predictions and optimizations for transferring circuit patterns onto silicon wafers to create integrated circuits [1]. As devices become increasingly miniaturized and designs grow more intricate, achieving high-resolution patterns becomes increasingly challenging. This amplifies the role of lithography simulation in electronic design automation (EDA), where it accelerates design cycles, integrates manufacturability considerations early, and ensures adherence to strict performance and yield standards. By minimizing the cost and time of experimental iterations, lithography simulation has become essential for addressing the complexities of modern semiconductor technologies and advancing the efficiency and accuracy of the entire EDA workflow.

Predictive simulation of lithographic processes requires the modeling of various physical and chemical phenomena. Typical lithography simulation flow for chemically amplified resist (CAR) consists of optical simulation and photoresist simulation, as can be seen in Fig. 1. Optical simulation models the light exposure process where light interacts with photomasks to project patterns onto the photoresist layer. In recent years, significant advancements have been made in optical simulation, driven by the need for precise control over light exposure and pattern fidelity [2]–[8]. Advanced techniques, such as source-mask optimization (SMO), further enhance optical simulation by fine-tuning exposure parameters [9], [10]. Photoresist simulation, on the other aspect, models the chemical and physical processes occurring within the photoresist layer during lithography, from exposure through post-exposure bake (PEB) to development. PEB is crucial in lithography, as it mitigates the standing wave effect and ensures consistent feature development [11]. In positive-tone CAR, exposure to an aerial image decomposes the photoacid generator, creating an acid latent image within the resist. During the subsequent PEB process, the resist is baked at elevated temperature, triggering an acid-catalyzed deprotection reaction that decomposes the

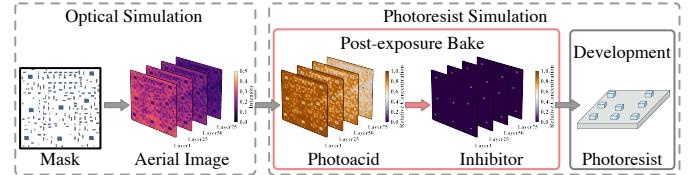


Fig. 1 A typical flow of lithography simulation for chemically amplified resist: from optical simulation to photoresist simulation.

inhibitor material [12]. By the end of PEB, the resist in the exposed area is deprotected, resulting in a higher development rate compared to unexposed area. The inhibitor concentration distribution determines the final resist profile after development. In regions where inhibitor concentration decreases due to acid reactions, the resist dissolves, forming the desired pattern. PEB simulation simulates the remaining inhibitor concentration based on the initial 3D photoacid profile, working as a connection between optical simulation and final printed features. Fine-tuning simulation tools is essential to accurately model the acid's catalytic behavior and predict final feature sizes.

According to the studies in [13], PEB simulation accounts for up to 30% of the runtime in the rigorous lithography simulation flow of Synopsys Sentaurus Lithography (S-Litho), which is non-negligible. Compared to machine learning-based optical simulation, PEB simulation often requires several times more computational time, highlighting the need for more efficient modeling approaches. Early research from the 1990s to early 2000s focused on acid-catalyzed reactions and physical dynamics governing pattern formation, using simplified reaction-diffusion equations [14], [15], and 3D diffusion profile simulations [16]. These methods demand significant computational resources, especially when simulating large-scale patterns. Methods such as finite element analysis (FEA) and finite difference methods (FDM) achieve spatial and temporal precision but are computationally intensive and limited in scalability for large-scale patterns. The lack of adaptive mesh refinement and seamless integration into lithography workflows further hinders efficiency and accuracy. Despite advancements in optical simulation, PEB simulation, crucial for accurate patterning, remains comparatively underexplored and could benefit from further development. Unlike 2D simulations, the 3D distribution characteristic of PEB introduces significant challenges in modeling complexity and computational cost. Recent state-of-the-art work DeePEB [13] addresses PEB modeling in CAR using neural networks, leveraging Fourier Neural Operator (FNO) [17] and convolution network to capture low-frequency global information and high-frequency local details. However, this approach struggles to fully capture continuous spatial and depthwise dependencies in 3D space, which are critical in PEB simulation where gradients and interactions across all three dimensions influence outcomes. Additionally, the frequency segmentation in FNO risks incomplete representations, while convolution operations may miss intricate high-frequency details required in 3D structures. The computational overhead of FNO

also impacts its efficiency, posing further challenges.

To address the above challenges in precisely and efficiently modeling PEB process, we propose SDM-PEB, designed to fully capture the spatial and depthwise dependencies inherent in complex physical and chemical reactions. Inspired by the recent advancements in Vision Transformers (ViT) [18]–[20], we adopt hierarchical transformer-based encoders with efficient spatial self-attention to extract both high-resolution coarse features and low-resolution fine-grained sequence features. Given that neighboring layers in the photoacid distribution provide valuable context, spatial and depthwise correlations require a carefully designed self-attention mechanism. Mamba [21], [22], a recent innovation based on state space models, efficiently captures long-range dependencies through selective scanning of feature sequences. Building on these new advancements, we introduce a customized spatial-depthwise Mamba-based attention unit (SDM unit) in each encoder layer, integrating parallel three-direction selective scan mechanisms of space state sequences to address both intra-layer spatial coherence and inter-layer interactions. Furthermore, a novel PEB focal loss is proposed to address the issue of imbalanced distribution of inhibitor. A differential divergence regularization is utilized to account for inter-layer variations. Our major contributions are summarized as follows:

- We employ a hierarchical vision transformer encoder to extract multi-scale spatial information within a single photoacid depth level.
- We propose a customized spatial-depthwise Mamba-based attention unit to enhance the model's capability in capturing cross-depth-level dependencies.
- We proposed the PEB focal loss and the differential depthwise divergence regularization to refine the learning objective, addressing reaction imbalances and reducing depthwise divergence.
- We evaluate our framework using the industry-verified software S-Litho, and the results demonstrate superior precision and efficiency compared to existing methods.

II. PRELIMINARIES

A. Photoresist Simulation Process

Post-Exposure Bake Process. Different resists undergo distinct physical and chemical reactions during PEB [11]. This paper focuses on positive-tone chemically amplified resists (CAR), the predominant materials used in ArF (193 nm) exposure for advanced semiconductor manufacturing. Modern CARs comprise protected polymers or dissolution inhibitors (\mathcal{I}), photoacid generators (PAG), and quencher bases (\mathcal{B}) [23]. The reaction process aims to create solubility contrast in the resist, enabling pattern formation. During lithographic exposure prior to PEB, incident light decomposes the PAG, producing photoacid (\mathcal{A}). In the subsequent thermal-driven PEB process, the photoacid catalyzes a reaction that deblocks the inhibitor. This catalytic reaction amplifies the effects of exposure and can be described by the catalytic reaction equation:

$$\frac{\partial[\mathcal{I}]}{\partial t} = -k_c[\mathcal{I}][\mathcal{A}], \quad (1)$$

where $[\mathcal{I}]$ and $[\mathcal{A}]$ denote the normalized concentration distributions of the inhibitor and photoacid, respectively, and k_c represents the catalysis coefficient.

The generated photoacid molecules diffuse within the resist during the process and are neutralized by the base quencher upon contact. This acid-base neutralization regulates acid diffusion, preventing excessive spreading and reactions in undesired regions. The base quencher also diffuses during PEB. The general reaction-diffusion

process for normalized concentration distribution of photoacid $[\mathcal{A}]$ and base quencher $[\mathcal{B}]$ can be expressed:

$$\frac{\partial[\mathcal{A}]}{\partial t} = -k_r[\mathcal{A}][\mathcal{B}] + D_{\mathcal{A}}\nabla^2[\mathcal{A}], \quad (2)$$

$$\frac{\partial[\mathcal{B}]}{\partial t} = -k_r[\mathcal{A}][\mathcal{B}] + D_{\mathcal{B}}\nabla^2[\mathcal{B}]. \quad (3)$$

In the above equation, the first term represents the neutralization reaction with k_r the reaction coefficient, the second term is Fick's law of diffusion in three dimensions, $D_{\mathcal{A}}$, $D_{\mathcal{B}}$ are the diffusion coefficients. The diffusion length L depends on diffusion coefficient and PEB duration T : $L = \sqrt{2D \cdot T}$.

For initial conditions (I.C.), without loss of generality, we consider uniform base and inhibitor distributions, while photoacid concentration is derived from the 3D aerial image via the Dill Model [24]. Boundary conditions (B.C.) are critical in confined structures like thin films. In x - y dimensions, we consider a commonly used zero-flux B.C., ensuring no flux across boundaries [25], [26]. In the z dimension, Robin B.C. describes photoacid in/out-diffusion at the resist surface:

$$\frac{\partial[\mathcal{A}]}{\partial n} = 0, \quad D_{\mathcal{A}} \frac{\partial[\mathcal{A}]}{\partial z} = h([\mathcal{A}]_{top} - [\mathcal{A}]_{sat}), \quad (4)$$

where n is the normal direction, h is the transfer coefficient and $[\mathcal{A}]_{sat}$ is the saturation concentration.

Development Process. After PEB, the development process transforms the chemically modified photoresist into a physical pattern. During this stage, exposed regions dissolve in the developer solution at a rate R , which depends on the inhibitor concentration and is described by the Mack kinetic model [27]:

$$R(x, y, z) = R_{max} \frac{(a+1)(1 - [\mathcal{I}])^n}{a + (1 - [\mathcal{I}])^n} + R_{min}, \quad a = (1 - M_{th})^n \frac{n+1}{n-1}. \quad (5)$$

Here R_{max} and R_{min} are the maximum and minimum development rates for fully exposed and unexposed regions, respectively. n is the surface reaction order, M_{th} is the experimentally determined threshold.

Using local development rates, the resist profile at a given development time t_d can be evaluated. The propagation of the development front $S_d(x, y, z)$ is governed by the Eikonal equation [28]: $|\nabla S_d(x, y, z)| = 1/R(x, y, z)$. In this work, we compute development rates from the inhibitor distribution and use an open-source Eikonal Solver [29] to calculate the development profile.

B. State Space Models-based Methodologies

Feature representations in 3D photoresist, characterized by consecutive depthwise distributions, can be naturally modeled as sequences of depth-levels (from shallow to deep). Recent advances in state space models (SSMs) have demonstrated their effectiveness in capturing long-range dependencies while supporting efficient parallel training. Originating from the Kalman filter [30], SSMs are linear time-invariant systems that map a scalar sequence $x(t)$ to another scalar sequence $y(t)$ via a hidden state $\mathbf{h}(t) \in \mathbb{R}^N$, where $t \in \{1 \dots L\}$. The system evolves with evolution parameter $\mathbf{A} \in \mathbb{R}^{N \times N}$, initialized using HiPPO matrix [31], and the projection parameters $\mathbf{B} \in \mathbb{R}^{N \times 1}$ and $\mathbf{C} \in \mathbb{R}^{N \times 1}$ as:

$$\mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}x(t), \quad y(t) = \mathbf{C}\mathbf{h}(t). \quad (6)$$

To adapt to deep learning, continuous-time SSMs are discretized using the zero-order hold (ZOH) assumption [32], transforming continuous-time parameters \mathbf{A} and \mathbf{B} into discrete equivalents:

$$\bar{\mathbf{A}} = \exp(\Delta \mathbf{A}), \quad \bar{\mathbf{B}} = (\Delta \mathbf{A})^{-1}(\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}, \quad (7)$$

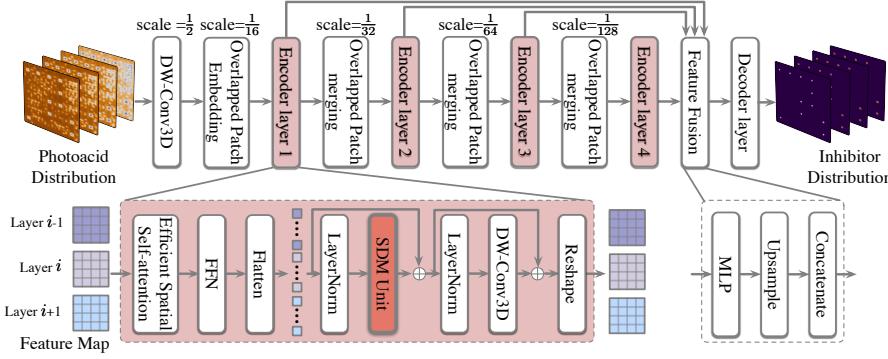


Fig. 2 The architecture overview of our proposed SDM-PEB framework.

with a specified time step Δ . The Equation (6) can be expressed in the discrete form:

$$\mathbf{h}_t = \bar{\mathbf{A}}\mathbf{h}_{t-1} + \bar{\mathbf{B}}x_t, \quad y_t = \mathbf{C}\mathbf{h}_t. \quad (8)$$

To improve computational efficiency and scalability, the iterative process in Equation (8) is reformulated as a single global convolution:

$$\mathbf{y} = \mathbf{x} * \bar{\mathbf{K}}, \quad \bar{\mathbf{K}} = (\mathbf{C}\bar{\mathbf{B}}, \mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}, \dots, \mathbf{C}\bar{\mathbf{A}}^{L-1}\bar{\mathbf{B}}), \quad (9)$$

where $*$ denotes the convolution operation.

Mamba [21], a recent selective scan SSM, excels in long-sequence modeling with linear complexity. By associating SSM parameters with the input, Mamba selectively focuses on relevant information while ignoring irrelevant inputs. This is achieved by configuring the projection matrix to be dependent on input sequence \mathbf{x} with K channels:

$$\mathbf{B} = \text{Linear}_N(\mathbf{x}), \quad \mathbf{C} = \text{Linear}_N(\mathbf{x}), \quad (10)$$

$$\Delta = \text{softplus}(\text{Broadcast}_K(\text{Linear}_1(\mathbf{x})) + \mathbf{D}), \quad (11)$$

where Linear_N projects to a hidden dimension N , Broadcast_K expands 1D data into a K -dimensions, and \mathbf{D} is a constant weight matrix. To enhance GPU utilization and memory efficiency, Mamba employs a hardware-aware algorithm for selective SSM computation with linear scalability relative to sequence length. By combining kernel fusion and recomputation within parallel scans, this approach minimizes memory I/O operations, resulting in substantial improvements in computation speed.

C. Problem Formulation

To enhance PEB simulation accuracy, we focus on predicting the 3D inhibitor distribution, known as the PEB latent image. Following prior work [13], we evaluate the predicted inhibitor distribution \mathcal{I} , development rate R , and resulting photoresist profile using the following metrics:

Root Mean Squared Error (RMSE). Root Mean Squared Error is commonly used to measure the difference between the predicted $\hat{\mathbf{P}}$ and the ground truth \mathbf{P} :

$$\text{RMSE} = \sqrt{\frac{\|\hat{\mathbf{P}} - \mathbf{P}\|^2}{n}}, \quad \mathbf{P} = [\mathcal{I}], \quad \mathbf{R}. \quad (12)$$

Normalized Root Mean Squared Error (NRMSE). Normalized Root Mean Squared Error (NRMSE) accounts for the magnitude of the true values, making the error more interpretable, especially for datasets with different scales or units:

$$\text{NRMSE} = \frac{\|\hat{\mathbf{P}} - \mathbf{P}\|_F}{\|\mathbf{P}\|_F}, \quad \mathbf{P} = [\mathcal{I}], \quad \mathbf{R}. \quad (13)$$

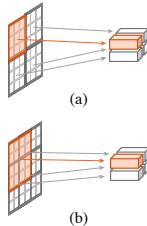


Fig. 3 (a) Non-overlapped patch merging and (b) overlapped patch merging.

where $\|\cdot\|_F$ denotes the Frobenius norm.

Photoresist Profile Critical Dimension Error (CD Error). In addition to the direct comparison of inhibitor distribution and development rate, CD error assesses the photoresist profile's accuracy by comparing printed feature dimensions CD_d in the x - y plane to target values:

$$\text{CD Error}_d = \sqrt{\frac{(\hat{CD}_d - CD_d)^2}{n}}, \quad d = x, y. \quad (14)$$

Problem 1 (PEB simulation problem). Given initial 3D photoacid distributions $\{\mathcal{A}_i\}_{i=1}^N$ and corresponding inhibitor distributions $\{\mathcal{I}_i\}_{i=1}^N$ with spatial dimensions D, H, W (depth, height, width), the objective is to develop and train a deep-learning model $\mathcal{F}_\Theta : [\mathcal{A}] \mapsto [\mathcal{I}]$, parameterized by Θ , to accurately predict inhibitor distributions for test mask patterns. The model must capture both spatial and depthwise dependencies, minimizing RMSE, NRMSE, and CD error while ensuring efficient computational runtime.

III. FRAMEWORK

A. Overview

In post-exposure bake (PEB) simulation, accurately modeling spatial (x - y plane) and depthwise (z -axis) dependencies is crucial for predicting complex 3D distributions. To enhance the model's ability to capture nuanced details essential for precise pattern representation in lithography, we propose the following:

- 1) **Hierarchical Contextual Feature Extractor**, designed to capture both coarse and fine-grained spatial features at each depth level;
- 2) **Spatial-Depthwise Mamba-based Attention Unit**, developed to model cross-depth-level dependencies effectively.

Fig. 2 provides an overview of the SDM-PEB architecture. The design of the proposed framework is detailed in this section.

B. Hierarchical Contextual Feature Extractor

A single-scale feature extractor with fixed resolution often struggles to effectively capture the complex patterns and multi-scale dependencies inherent in the PEB simulation task. Inspired by advancements in hierarchical feature extractors [33], [34], we adopt a hierarchical feature extractor to generate high-resolution coarse features and low-resolution fine-grained features. To further optimize the trade-off between performance and efficiency in the PEB simulation task, we incorporate two key customizations.

Depthwise Overlapped Patch Merging. Local dependencies between neighboring regions are essential for modeling reactions influenced by adjacent concentrations. These dependencies help the model understand photoacid diffusion and reaction rate variations.

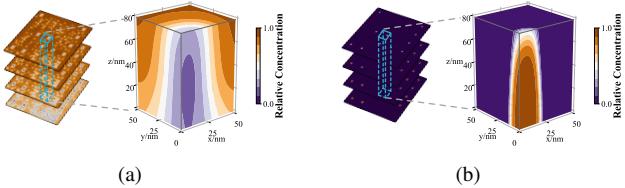


Fig. 4 Vertical visualization of distributions: (a) photoacid at the initial stage and (b) inhibitor at the final stage.

To reduce information loss at patch boundaries and enhance local continuity, we apply depthwise overlapped patch merging instead of non-overlapped patch merging to downsample spatial feature maps while retaining depth resolution. A comparative illustration is provided in Fig. 3. Additionally, The channel dimension increases after each merging layer, capturing increasingly complex and abstract feature representations.

Efficient Spatial Self-Attention. ViT-based feature extractors have achieved great success in general computer vision tasks by employing spatial self-attention to dynamically assign importance to input features. However, the computational complexity of standard multi-head self-attention is $O(L^2)$, where L is the sequence length of the input.

In the task of PEB prediction, the input photoacid distributions are typically large due to high-resolution requirements and include a depth dimension, making the computational cost of standard attention mechanisms infeasible. To address this challenge, we apply a reduction ratio r to shorten the sequence length, as suggested in [33]. The transformation is defined as follows:

$$\hat{\mathbf{K}} = \text{Reshape} \left(\frac{L}{r}, C \cdot r \right) (\mathbf{K}), \quad \mathbf{K} = \text{Linear}_C(\hat{\mathbf{K}}), \quad (15)$$

where \mathbf{K} is the key in the attention mechanism, and C is the feature dimension of \mathbf{K} . After the transformation, the sequence length of \mathbf{K} is reduced by the reduction ratio r , decreasing the computational complexity from $O(L^2)$ to $O(L^2/r)$, thereby making the computational cost manageable.

C. Spatial-Depthwise Mamba-based Attention Unit

As described in Equations (1) and (2), photoacid molecules generated during exposure diffuse and trigger chemical reactions in adjacent layers, altering the inhibitor profile. Therefore, depthwise interactions are crucial for accurately predicting the final inhibitor distribution and resist profile. This causal depthwise dependency can be observed in the vertical profiles shown in Fig. 4, which exhibit continuous and gradual changes in photoacid and inhibitor distributions.

Given that Mamba [21] is renowned for its ability to model sequential information, we propose the spatial-depthwise Mamba-based attention unit (SDM Unit) to extract inter-depth-level relationships. The SDM Unit extends bidirectional scanning into three dimensions to better aggregate depthwise information and employs structured state space models to effectively capture both spatial and depthwise dependencies.

Spatial-Depthwise PEB Selective Scan. As introduced in Section II, Mamba is traditionally designed for 1D sequences. To extend Mamba to the PEB simulation task and enhance its capability for 3D photoacid-inhibitor modeling, the selective scan path must be carefully designed. Building on the ideas from previous works [22], [35], [36], we propose employing three types of selective scans simultaneously: (1) **Spatial Scan**: The spatial scan operates along the depth dimension, focusing on information at a specific spatial position

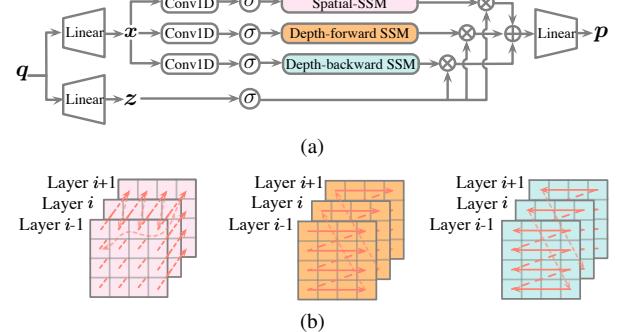


Fig. 5 (a) The architecture of the spatial-depthwise Mamba-based attention unit; (b) Illustration of the three-direction PEB selective scan, from left to right: spatial scan, depth-forward scan, and depth-backward scan.

across all depth layers; (2) **Depth-Forward Scan**: The depth-forward scan processes the entire shallow level first before transitioning to deeper levels; (3) **Depth-Backward Scan**: The depth-backward scan follows the reverse order of the depth-forward scan, processing deeper levels before moving to shallower ones. We illustrate the scan methods in Fig. 5(b). The spatial-depthwise three-direction PEB selective scan, serves as the core element in the SDM Unit, and leverages parallel SSMS to establish the intra- and inter-depth long-range dependencies.

Structure of SDM Unit. The structure of the spatial-depthwise Mamba-based attention unit (SDM Unit) is illustrated in Fig. 5(a). Specifically, the feed-forward layer output in the i -th encoder layer, comprising feature map embeddings with dimensions $\mathbb{R}^{C_i \times D \times H_i \times W_i}$, is reshaped into a sequence \mathbf{q}_i of size $\mathbb{R}^{C_i \times DH_i W_i}$, and normalized through a normalization layer. This sequence is then fed into the SDM Unit, which enhances both inter-layer and intra-layer distribution dependencies. The normalized sequence \mathbf{q}_i is linearly projected into \mathbf{x}_i and \mathbf{z}_i with hidden dimension C_i^h . For each scanning direction, a 1D convolution followed by SiLU activation [37] is applied to \mathbf{x} , resulting in \mathbf{x}'_d as input for the d -direction spatial-depthwise PEB selective scan: $\mathbf{B}, \mathbf{C}, \Delta$ are derived from Equation (10), and discretized values $\bar{\mathbf{A}}, \bar{\mathbf{B}}$ are computed using Equation (7). The outputs of the selective scan are calculated via Equation (9). These outputs are then weighted by \mathbf{z}_i and combined to produce the final feature \mathbf{p}_i . A depthwise convolution with a kernel size of 3 is subsequently applied to refine fine-grained features.

By dynamically adjusting attention weights across spatial and depthwise dimensions, the model effectively captures both local and global patterns within the resist layers. This enables accurate modeling of intricate physical-chemical interactions in 3D inhibitor distributions, resulting in a precise inhibitor profile.

D. Customized PEB Optimization Objectives

To accommodate the exponential variation of inhibitor as indicated in Equation (1), we apply the label normalization technique from [13], specifically predicting the quadratic negative logarithmic transformation of the inhibitor distribution as the output \mathcal{Y} rather than the raw distribution. This transformation is defined as $\mathcal{Y} = -\ln(-\ln([\mathcal{I}])/k_c)$. In previous work [13], the authors used the single maximum squared error (MaxSE):

$$\mathcal{L}_{\text{MaxSE}} = \max_{d,h,w} \left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w} \right)^2. \quad (16)$$

However, by emphasizing only the maximum error, MaxSE may not reflect the overall quality of predictions across the entire distribution. Additionally, the model could become unstable, potentially oscillating

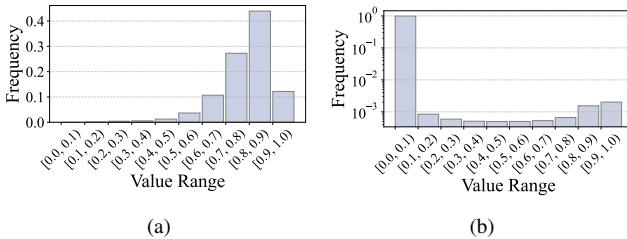


Fig. 6 Distribution frequencies of (a) photoacid value ranges and (b) inhibitor value ranges.

to correct extreme outliers rather than achieving balanced accuracy. In this work, we introduce a customized PEB focal loss and a differential depth divergence regularization to address the complexity of the 3D distribution.

PEB Focal Loss. As can be seen from Fig. 6, the distributions of both photoacid and inhibitor are highly imbalanced: In every layer, photoacid spreads widely but shows significant exception inside contacts meanwhile inhibitor tends to be localized inside contacts. The frequency of the inhibitor distribution can even differ by several orders of magnitude. This imbalance skews the data toward low-variance background regions, making it difficult for the model to focus on subtle but critical changes near contacts. Inspired by the focal loss proposed for hard example mining [38], we design a PEB focal loss to address the imbalance issue:

$$\mathcal{L}_{\text{PEB-FL}} = \sum_d^D \sum_h^H \sum_w^W \left| \hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w} \right|^{\gamma} \left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w} \right)^2, \quad (17)$$

where γ is the focusing parameter, empirically set to 1 in our experiment.

Differential Depth Divergence Regularization. In 3D PEB simulation, understanding how the inhibitor distribution evolves across layers is crucial for enhancing spatial-depthwise dependencies. We further propose the differential depth divergence regularization focusing on aligning inter-layer differences. For every pair $\hat{\mathcal{Y}}, \mathcal{Y} \in \mathbb{R}^{D \times H \times W}$, we first calculate their layer-wise forward difference maps $\Delta\hat{\mathcal{Y}}, \Delta\mathcal{Y} \in \mathbb{R}^{(D-1) \times H \times W}$, in which d -th layer can be expressed:

$$\Delta\hat{\mathcal{Y}}_d = \hat{\mathcal{Y}}_{d+1} - \hat{\mathcal{Y}}_d, \quad \Delta\mathcal{Y}_d = \mathcal{Y}_{d+1} - \mathcal{Y}_d. \quad (18)$$

We then convert the difference maps into probabilities by applying the softmax function across the height and width dimensions, in which high difference layers are penalized:

$$\sigma(\Delta\hat{\mathcal{Y}}_d) = \frac{\exp(\Delta\hat{\mathcal{Y}}_d/\tau)}{\sum_{h=1}^H \sum_{w=1}^W \exp(\Delta\hat{\mathcal{Y}}_{d,h,w}/\tau)}, \quad (19)$$

$$\sigma(\Delta\mathcal{Y}_d) = \frac{\exp(\Delta\mathcal{Y}_d/\tau)}{\sum_{h=1}^H \sum_{w=1}^W \exp(\Delta\mathcal{Y}_{d,h,w}/\tau)}, \quad (20)$$

where τ is the parameter to sharpen the probability distribution, we empirically set to 0.1 here. The differential depth divergence regularization \mathcal{L}_{Div} is formulated as the Kullback-Leibler divergence between the predicted difference map and the ground difference map:

$$\mathcal{L}_{\text{Div}} = \sum_{d=1}^{D-1} \sigma(\Delta\hat{\mathcal{Y}}_d) \log \frac{\sigma(\Delta\hat{\mathcal{Y}}_d)}{\sigma(\Delta\mathcal{Y}_d)}. \quad (21)$$

The overall loss function is the linear combination of the three weighted by α and β with empirical values 1.0 and 0.1. The learning on the whole distribution can be enhanced:

$$\mathcal{L} = \mathcal{L}_{\text{MaxSE}} + \alpha \mathcal{L}_{\text{PEB-FL}} + \beta \mathcal{L}_{\text{Div}}. \quad (22)$$

TABLE I Important parameters in photoresist simulation process.

PEB			
Normal Diffusion Length $L_{N,A}, L_{N,B}$	70, 15 nm	Lateral Diffusion Length $L_{L,A}, L_{L,B}$	10, 10 nm
catalysis coefficient k_c	0.9 /s	reaction coefficient k_r	8.6993 /s
transfer coefficient h_A, h_B	0.027, 0	saturation concentration $[\mathcal{A}]_{\text{sat}}, [\mathcal{B}]_{\text{sat}}$	0.9, 0
$ \mathcal{T} (t = 0)$	1.0	$ \mathcal{B} (t = 0)$	0.4
Baseline Time step	0.1 s	Duration	90 s
Develop			
R_{max}	40 nm/s	R_{min}	0.0003 nm/s
M_{th}	0.5	n	30
Duration	60 s		

TABLE II Comparison with different PEB solvers.

Methodologies	Inhibitor		Develop Rate		CD Error x (nm)	y (nm)	RT/s
	RMSE (e-3)	NRMSE (%)	RMSE (nm/s)	NRMSE (%)			
DeepCNN [39]	8.25	12.53	0.65	1.63	3.14	6.26	1.01
TEMPO-resist [3]	7.67	12.55	0.50	1.26	2.12	2.45	6.48
FNO [17]	7.91	11.68	0.68	1.69	2.34	3.71	1.15
DeePEB [13]	3.99	5.70	0.48	1.19	0.98	1.24	1.37
SDM-PEB	2.78	3.70	0.35	0.86	0.74	0.93	1.06

IV. EXPERIMENTS

We generate 100 mask clips from [40], each sized at $2 \times 2 \mu\text{m}^2$, which proved sufficient in our experiments. These mask clips are designed with contact sizes and distribution patterns suitable for technology nodes at 28nm and below. Rigorous simulations are conducted using S-Litho. The wavelength is $\lambda = 193\text{nm}$ and numerical aperture is $NA = 1.35$ for the optical exposure simulation, the optical influence range of $5\lambda/NA$ and simulation resolution of 0.5nm in three directions are configured for the $2 \times 2 \mu\text{m}^2$ simulation window. For the PEB simulation and development simulation, the resist thickness is 80nm , and resolutions in x , y and z directions are 2nm , 2nm , and 1nm respectively. Other important simulation parameters included are listed in TABLE I.

In the SDM-PEB configuration, the spatial sizes of feature maps from encoder layers 1 to 4 are reduced to $[\frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}]$ of raw input, achieved using patch sizes of $[15, 3, 3, 3]$ and strides of $[8, 2, 2, 2]$. The feature dimensions for these layers are $[64, 128, 320, 512]$, respectively. Self-attention reduction ratios are $[64, 16, 4, 1]$. The MLP in the feature fusion layer has 768 dimensions. The decoder includes 3 transpose convolution layers, with LeakyReLU activations between them. The training was conducted on 2 NVIDIA RTX 3090 GPUs for 500 epochs using a step decay scheduler, beginning at a learning rate of 0.03 with a step size of 100 and a decay factor of 0.7. We used a batch size of 8 by accumulating gradients over 8 clips before updating the model each epoch, considering the restriction of GPU memory. For a fair comparison, we followed the same train-test split as prior methods and tested runtime on our machines.

Compare With Learning-based PEB solvers. There exists few prior works focused on modeling photoresist simulation. We compared our SDM-PEB with previous state-of-the-art (SOTA) learning-based photoresist solvers. DeepCNN is customized from [39] with a residual connection for adaption to our problem. TEMPO-resist is modified from TEMPO [3], which originally used a conditional-GAN to predict 3D aerial images at various heights, to suit our 3D PEB simulation. DeePEB [13] extends FNO [17] by integrating CNN-based local learning branches to capture high-frequency information.

As is shown in TABLE II, our SDM-PEB outperforms all compared learning-based methods by a significant margin. Compared to the previous best method, DeePEB, SDM-PEB achieves 43.5% lower average RMSE and 54.1% lower NRMSE for inhibitor concentration predictions. For development rate prediction, with values bounded between R_{min} and R_{max} , our method reduces RMSE by 37.1% and

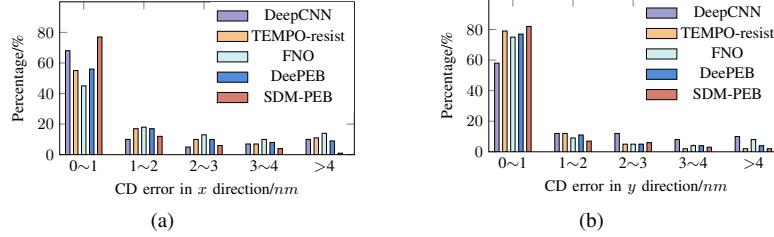


Fig. 7 Percentage counts of CD errors using different methods: (a) error in the x direction and (b) error in the y direction.

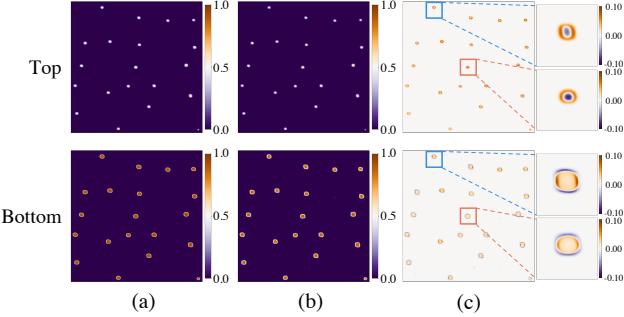


Fig. 8 Top-down visualization examples of predicted distribution results. The upper row is the top surface and the lower row is the bottom surface. (a) Ground truths, (b) predictions and (c) differences.

NRMSE by 38.4%. This superior predictive accuracy extends to the final resist profile, improving CD error by 32.4% in the x -direction and 33.3% in the y -direction. Compared to DeepCNN, TEMPO-resist, and FNO, the accuracy improvements are even more pronounced across all evaluated metrics.

The distributions of CD errors in the x and y directions are shown in Fig. 7. Compared to other methods, the CD errors of SDM-PEB are more concentrated within the range of 0–1 nm, demonstrating superior prediction accuracy. Furthermore, the CD error distributions in the x and y directions are more consistent, demonstrating the robustness and stability of our approach.

SDM-PEB also demonstrates high efficiency, achieving an average runtime (RT) of 1.06s, which is 138× faster than S-Litho's 147s. It achieves a 6× speedup compared to TEMPO-resist and reduces runtime by 8.5% and 29.2% compared to FNO and DeePEB, respectively. Although it takes 4.7% more time than DeepCNN, this slight runtime overhead is justified by its substantial performance improvements.

Visualization of Simulation Results. Fig. 8 presents examples comparing the ground truth with SDM-PEB results at both the top layer and bottom surface. The predicted inhibitor shows minimal deviation from the ground truth, with absolute errors across most positions on the plane remaining within 0.1. Spatial information is effectively extracted and translated, as evidenced by the consistency of results at both the center and the corner. Fig. 9 compares the vertical profiles of contacts at the center and corner, corresponding to contacts within the red box and blue box in Fig. 8. Leveraging the three-direction PEB selective scan for depthwise information gathering, SDM-PEB successfully captures relationships between adjacent depth distributions and models causal variations across different depths, achieving consistent simulations along the depth direction. The primary discrepancies occur at the edges of contacts, where concentration changes are more drastic. At the top layer, the inhibitor concentration is lower than at the bottom and is confined to a smaller

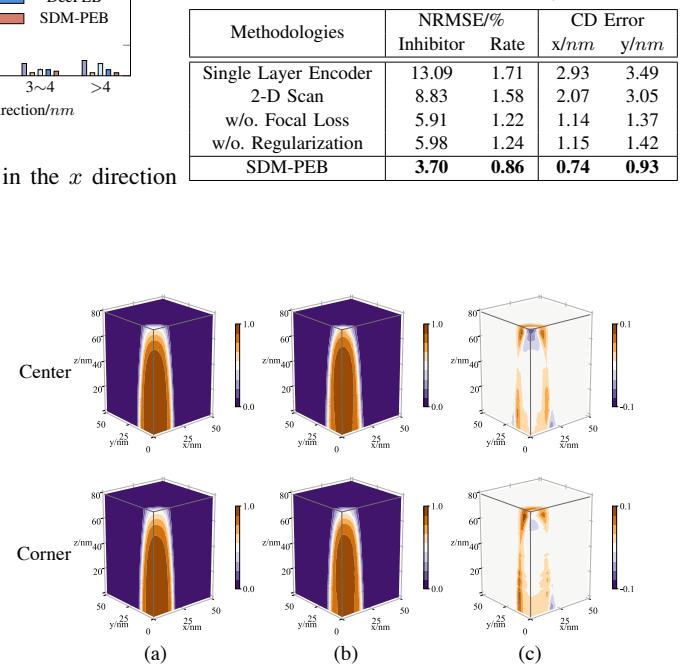


Fig. 9 Vertical visualization of predicted results: the upper row shows the center contact, the lower row shows the corner contact. (a) Ground truths, (b) predictions, (c) differences.

region, adding to the challenges for the model.

Ablation Study. We conducted an ablation study on each component of our method, with the results presented in TABLE III. In the Single Layer Encoder setting, only the feature map from the first encoder layer is used, which fails to capture complex features and results in the lowest accuracy. In the 2-D scan setting, only the depth-forward and depth-backward scans shown in Fig. 5(b) are utilized, which is adapted from [22] for our 3D tasks. Compared to this setting, SDM-PEB achieves significant improvements across all metrics, demonstrating the importance of depth dependencies in enhancing PEB modeling. Furthermore, we remove the PEB focal loss and differential depth divergence regularization in the ‘w/o. Focal Loss’ and ‘w/o. Regularization’, respectively, while retaining the other two optimization objectives. The results show noticeable degradation compared to the full SDM-PEB, highlighting the critical role of each component in refining the framework.

V. CONCLUSION

This paper introduces SDM-PEB, a novel modeling framework designed to enhance 3D PEB simulation. It effectively captures spatial and depthwise dependencies to model the physical and chemical interactions underlying the inhibitor distributions while maintaining efficiency. Our framework integrates hierarchical feature extraction with efficient self-attention and spatial-depthwise Mamba-based attention to capture long-range dependencies. To address data imbalances and improve sensitivity to 3D variations, we propose a PEB focal loss and differential depth divergence regularization. Experimental results, validated with an industry-grade software S-Litho under real manufacturing conditions demonstrate the superior performance of our method. Through the development of modeling advanced photoresist simulation, we aim to further progress lithography techniques.

TABLE III Ablation study

Methodologies	NRMSE/%		CD Error	
	Inhibitor Rate	x/nm y/nm	x/nm	y/nm
Single Layer Encoder	13.09	1.71	2.93	3.49
2-D Scan	8.83	1.58	2.07	3.05
w/o. Focal Loss	5.91	1.22	1.14	1.37
w/o. Regularization	5.98	1.24	1.15	1.42
SDM-PEB	3.70	0.86	0.74	0.93

ACKNOWLEDGEMENTS

The project is supported in part by Research Grants Council of Hong Kong SAR (No. RFS2425-4S02, No. CUHK14211824 and No. CUHK14210723), and the MIND project (MIDXZ202404).

REFERENCES

- [1] C. A. Mack, "Thirty years of lithography simulation," in *Optical Microlithography XVIII*, vol. 5754. SPIE, 2005, pp. 1–12.
- [2] W. Ye, M. B. Alawieh, Y. Lin, and D. Z. Pan, "Lithogan: End-to-end lithography modeling with generative adversarial networks," in *Proc. DAC*, 2019, pp. 1–6.
- [3] W. Ye, M. B. Alawieh, Y. Watanabe, S. Nojima, Y. Lin, and D. Z. Pan, "Tempo: Fast mask topography effect modeling with deep learning," in *Proceedings of the 2020 International Symposium on Physical Design*, 2020, pp. 127–134.
- [4] Z. Yu, G. Chen, Y. Ma, and B. Yu, "A gpu-enabled level-set method for mask optimization," *IEEE TCAD*, 2022.
- [5] Z. Wang, Y. Shen, X. Yao, W. Zhao, Y. Bai, F. Farnia, and B. Yu, "Chatpattern: Layout pattern customization via natural language," in *Proceedings of the 61st ACM/IEEE Design Automation Conference*, 2024, pp. 1–6.
- [6] Z. Yu, P. Liao, Y. Ma, B. Yu, and M. D. Wong, "CTM-SRAF: Continuous transmission mask-based constraint-aware sub resolution assist feature generation," *IEEE TCAD*, 2023.
- [7] H. Yang, Z. Li, K. Sastry, S. Mukhopadhyay, M. Kilgard, A. Anandkumar, B. Khailany, V. Singh, and H. Ren, "Generic lithography modeling with dual-band optics-inspired neural networks," in *Proceedings of the 59th ACM/IEEE Design Automation Conference*, 2022, pp. 973–978.
- [8] G. Chen, Z. Pei, H. Yang, Y. Ma, B. Yu, and M. Wong, "Physics-informed optical kernel regression using complex-valued neural fields," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2023, pp. 1–6.
- [9] H. Sun, J. Du, C. Jin, H. Quan, Y. Li, Y. Tang, J. Wang, S. Hu, and J. Liu, "Global optimisation of source and mask in inverse lithography via tabu search combined with genetic algorithm," *Optics Express*, vol. 30, no. 14, pp. 24 166–24 185, 2022.
- [10] G. Chen, Z. Wang, B. Yu, D. Z. Pan, and M. D. Wong, "Ultra-fast source mask optimization via conditional discrete diffusion," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2024.
- [11] A. Erdmann, T. Füthner, F. Shao, and P. Evanschitzky, "Lithography simulation: modeling techniques and selected applications," *Modeling Aspects in Optical Metrology II*, vol. 7390, pp. 13–29, 2009.
- [12] J. Byers *et al.*, "Characterization and modeling of a positive chemically amplified resist," in *Proceeding of SPIE*, vol. 2438, 1995, pp. 153–166.
- [13] Q. Wang, X. Gao, Y. Lin, R. Wang, and R. Huang, "DeepEB: A neural partial differential equation solver for post exposure baking simulation in lithography," in *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, 2022, pp. 1–9.
- [14] R. A. Ferguson, J. M. Hutchinson, C. A. Spence, and A. R. Neureuther, "Modeling and simulation of a deep-ultraviolet acid hardening resist," *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena*, vol. 8, no. 6, pp. 1423–1427, 1990.
- [15] M. Weiss, H. Binder, and R. Schwalm, "Modeling and simulation of a chemically amplified duv resist using the effective acid concept," *Microelectronic engineering*, vol. 27, no. 1-4, pp. 405–408, 1995.
- [16] E. Tomacruz, M. Zuniga, R. Guerrieri, A. Neureuther, and A. Sangiovanni-Vincentelli, "3-d diffusion models for chemically-amplified resists using massively parallel processors," in *Simulation of Semiconductor Devices and Processes: Vol. 5*. Springer, 1993, pp. 109–112.
- [17] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," *arXiv preprint arXiv:2010.08895*, 2020.
- [18] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [19] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.
- [20] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *International conference on machine learning*. PMLR, 2021, pp. 10 347–10 357.
- [21] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.
- [22] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:2401.09417*, 2024.
- [23] M. D. Stewart, K. Patterson, M. H. Somervell, and C. G. Willson, "Organic imaging materials: a view of the future," *Journal of Physical Organic Chemistry*, vol. 13, no. 12, pp. 767–774, 2000.
- [24] F. H. Dill, W. P. Hornberger, P. S. Hauge, and J. M. Shaw, "Characterization of positive photoresist," *IEEE Transactions on electron devices*, vol. 22, no. 7, pp. 445–452, 1975.
- [25] C. A. Mack, *Modeling solvent effects in optical lithography*. The University of Texas at Austin, 1998.
- [26] M. D. Smith and C. A. Mack, "Examination of a simplified reaction-diffusion model for post-exposure bake of chemically amplified resists," in *Advances in Resist Technology and Processing XVIII*, vol. 4345. SPIE, 2001, pp. 1022–1036.
- [27] C. A. Mack, M. J. Maslow, A. Sekiguchi, and R. A. Carpio, "New model for the effect of developer temperature on photoresist dissolution," in *Advances in Resist Technology and Processing XV*, vol. 3333. SPIE, 1998, pp. 1218–1231.
- [28] C. Mack, *Fundamental principles of optical lithography: the science of microfabrication*. John Wiley & Sons, 2007.
- [29] W.-K. Jeong and R. T. Whitaker, "A fast iterative method for eikonal equations," *SIAM Journal on Scientific Computing*, vol. 30, no. 5, pp. 2512–2534, 2008.
- [30] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.
- [31] A. Gu, T. Dao, S. Ermon, A. Rudra, and C. Ré, "Hippo: Recurrent memory with optimal polynomial projections," *Advances in neural information processing systems*, vol. 33, pp. 1474–1487, 2020.
- [32] A. Gu, I. Johnson, K. Goel, K. Saab, T. Dao, A. Rudra, and C. Ré, "Combining recurrent, convolutional, and continuous-time models with linear state space layers," *Advances in neural information processing systems*, vol. 34, pp. 572–585, 2021.
- [33] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 568–578.
- [34] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.
- [35] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "Vmamba: Visual state space model," 2024.
- [36] C. Yang, Z. Chen, M. Espinosa, L. Ericsson, Z. Wang, J. Liu, and E. J. Crowley, "Plainmamba: Improving non-hierarchical mamba in visual recognition," *arXiv preprint arXiv:2403.17695*, 2024.
- [37] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural networks*, vol. 107, pp. 3–11, 2018.
- [38] T.-Y. Ross and G. Dollár, "Focal loss for dense object detection," in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2980–2988.
- [39] Y. Watanabe, T. Kimura, T. Matsunawa, and S. Nojima, "Accurate lithography simulation model based on convolutional neural networks," in *Optical Microlithography XXX*, vol. 10147. SPIE, 2017, pp. 137–145.
- [40] Y. Lin, M. Li, Y. Watanabe, T. Kimura, T. Matsunawa, S. Nojima, and D. Z. Pan, "Data efficient lithography modeling with transfer learning and active data selection," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 38, no. 10, pp. 1900–1913, 2018.