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

Ziyang Yu[1], Peng Xu[1], Zixiao Wang[1], Binwu Zhu[2], Qipan Wang[3], Yibo Lin[3], Runsheng Wang[3], Bei Yu[1], Martin Wong[4]

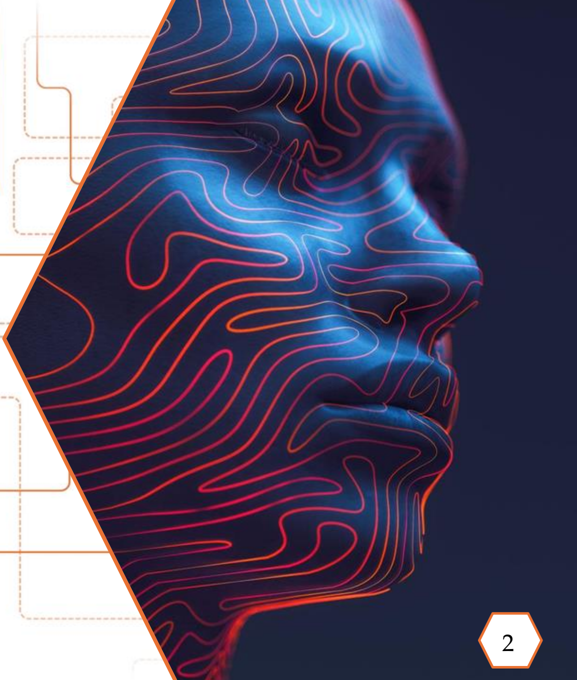
[1]The Chinese University of Hong Kong [2]Southeast University & NCTIEDA
[3]Peking University [4]Hong Kong Baptist University

THE CHIPS TO SYSTEMS CONFERENCE

SPONSORED BY CEDA sigda

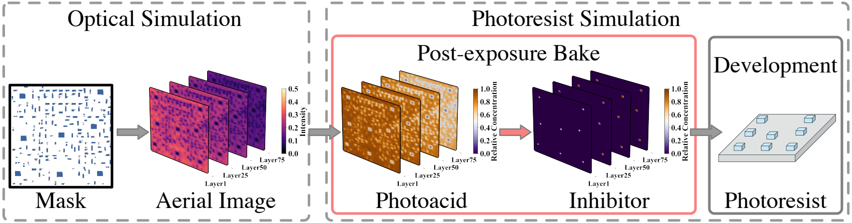
Outline

1. Introduction
2. Framework
3. Experiment



Introduction

Typical Lithography Simulation Flow



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

- **Optical simulation**: light exposure process
- **Photoresist simulation**: chemical and physical processes occurring within photoresist layer

Post Exposure Bake Process

Step 1: incident light decomposes photoacid generators, generating photoacid ($\mathcal{A}$).
Step 2: photoacid catalized inhibitor ($\mathcal{I}$) decomposition:

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

Step 3: photoacid-base quencher($\mathcal{B}$) neutralization & diffusion:

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

k_c: catalysis coefficient; k_r: reaction coefficient; $D_{\mathcal{A}}$, $D_{\mathcal{B}}$: the diffusion coefficients

Development Process

Step 4: photoresist developed at a rate $\boldsymbol{R}$:

$$\boldsymbol{R}(x,y,z) = R_{max}\frac{(a+1)(1-[\mathcal{I}])^n}{a+(1-[n])^n} + R_{min},\ a = (1-M_{th})^n\frac{n+1}{n-1}. \tag{4}$$

R_{max}, R_{min}: maximum (fully exposed) and minimum (unexposed) development rates;
n: surface reaction order

Importance of Improving PEB simulation

1. Accounts for 30% of the runtime in Synopsys Sentaurus Lithography (S-Litho)
2. Early Methods: significant computational burden
   - Simplified reaction-diffusion equations
   - 3D diffusion profile simulations
   - Finite element analysis
   - Finite difference methods
3. DeePEB-Fourier Neural Operator (FNO) + CNN: Fails to capture full 3D spatial-depth dependencies; information loss in frequency segmentation

Motivation: Fully capture the spatial and depthwise dependencies inherent in complex physical and chemical reactions.

Framework

THE CHIPS TO SYSTEMS CONFERENCE

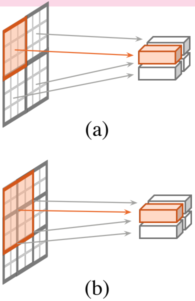
Our Contributions

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.
3. **Customized PEB Optimization Objectives**
   - efficiently guide the optimization.

Hierarchical Contextual Feature Extractor

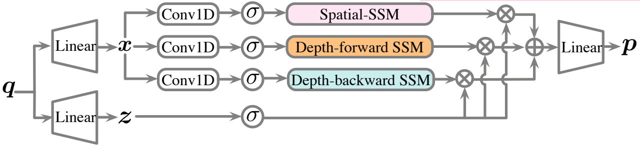
1. Depthwise Overlapped Patch Merging
   - Reduce information loss at patch boundaries
   - Enhance local continuity
2. Efficient Spatial Self-Attention:
   - C: feature dimension of $\boldsymbol{K}$; r: reduction ratio
   - Computational complexity: $O(L^2) \rightarrow O(L^2/r)$

$$\hat{\boldsymbol{K}} = \texttt{Reshape}\left(\frac{L}{r}, C \cdot r\right)(\boldsymbol{K}), \quad \boldsymbol{K} = \texttt{Linear}_C(\hat{\boldsymbol{K}}), \quad (5)$$



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

Spatial-Depthwise Mamba-based Attention Unit



The architecture of the spatial-depthwise Mamba-based attention unit.

1. Feature map with dimension $\mathbb{R}^{C_i \times D \times H_i \times W_i}$ reshaped into: $\boldsymbol{q}_i \in \mathbb{R}^{C_i \times DH_iW_i}$.
2. $\boldsymbol{q}_i$ linearly projected into $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ with hidden dimension C_i^h.
3. In each direction d: $\boldsymbol{x}_i \rightarrow$ 1D convolution $\rightarrow$ SiLU activation *rightarrow* d-direction spatial-depthwise PEB selective scan
4. Weighted and combined to produce feature map $\boldsymbol{p}$

Spatial-Depthwise PEB Selective Scan

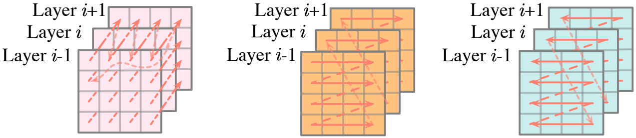


Illustration of the three-direction PEB selective scan, from left to right: spatial scan, depth-forward scan, and depthbackward scan.

- **Spatial Scan**: operates along the depth dimension to collect information at a specific spatial position across all depth layers
- **Depth-Forward Scan**: processes the entire shallow level first before transitioning to deeper levels
- **Depth-Backward Scan**: processes deeper levels before moving to shallower ones

State Space Model

State space model (SSM):

- Capture long-range dependencies with parallel training
- Map a scalar sequence $x(t)$ to another scalar sequence $y(t)$ via a hidden state $\boldsymbol{h}(t) \in \mathbb{R}^N$
- $\boldsymbol{A} \in \mathbb{R}^{N \times N}$: evolution parameter; $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{N \times 1}$: projection parameters

$$\boldsymbol{h}'(t) = \boldsymbol{A}\boldsymbol{h}(t) + \boldsymbol{B}x(t), \; y(t) = \boldsymbol{C}\boldsymbol{h}(t). \tag{6}$$

Deep learning adaption: zero-order hold (ZOH) discretization assumption:

$$\bar{\boldsymbol{A}} = \exp(\boldsymbol{\Delta A}), \; \bar{\boldsymbol{B}} = (\boldsymbol{\Delta A})^{-1}(\exp(\boldsymbol{\Delta A}) - \boldsymbol{I}) \cdot \boldsymbol{\Delta B}, \tag{7}$$

Discretized version re-expression:

$$\boldsymbol{h}_t = \bar{\boldsymbol{A}}\boldsymbol{h}_{t-1} + \bar{\boldsymbol{B}}x_t, \; y_t = \boldsymbol{C}\boldsymbol{h}_t. \tag{8}$$

Mamba: Selective Scan State Space Model

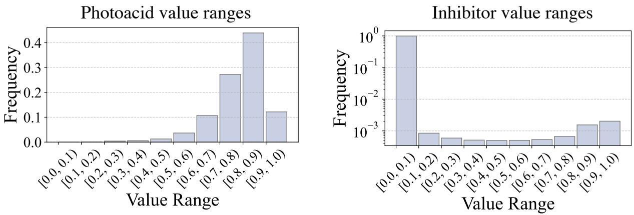
- Selectively focuses on relevant information while ignoring irrelevant inputs.
- Associate SSM projection parameters with the input
- Hardware-aware algorithm for SSM computation with linear scalability relative to sequence length
- Parallel scans: Kernel fusion and recomputation

$$\boldsymbol{B} = \texttt{Linear}_N(\boldsymbol{x}),\ \boldsymbol{C} = \texttt{Linear}_N(\boldsymbol{x}), \tag{9}$$

$$\boldsymbol{\Delta} = \texttt{softplus}(\texttt{Broadcast}_K(\texttt{Linear}_1(\boldsymbol{x})) + \boldsymbol{D}), \tag{10}$$

Customized PEB Optimization Objectives

- **Maximum squared error (MaxSE):** $\mathcal{L}_{\text{MaxSE}} = \max_{d,h,w} \left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right)^2$
- **PEB focal loss:**
  - distributions of both photoacid and inhibitor are highly imbalanced
  - $\mathcal{L}_{\text{PEB-FL}} = \sum_d^D \sum_h^H \sum_w^W \left|\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right|^{\gamma} \left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right)^2$



Customized PEB Optimization Objectives

- **Differential depth divergence regularization**: aligning inter-layer differences
  - For every pair $\hat{\mathcal{Y}}, \mathcal{Y} \in \mathbb{R}^{D\times H\times W}$, calculate layer-wise forward difference maps $\Delta\hat{\mathcal{Y}}, \Delta\mathcal{Y} \in \mathbb{R}^{(D-1)\times H\times W}$: $\Delta\hat{\mathcal{Y}}_d = \hat{\mathcal{Y}}_{d+1} - \hat{\mathcal{Y}}_d,\ \Delta\mathcal{Y}_d = \mathcal{Y}_{d+1} - \mathcal{Y}_d$
  - convert the difference maps into probabilities to penalize high difference layers:

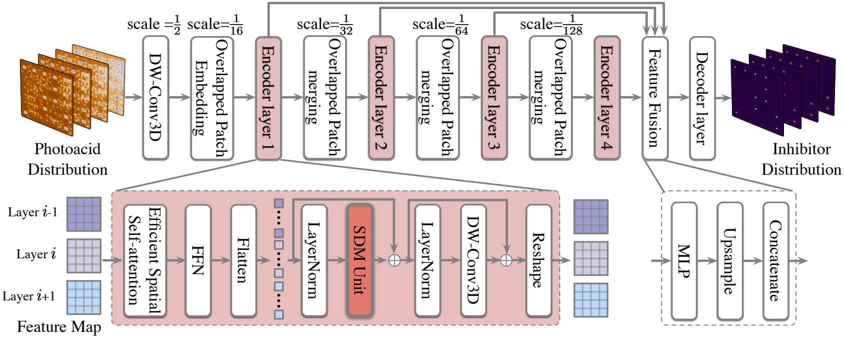
$$\sigma(\Delta\hat{\mathcal{Y}}_d) = \frac{\exp(\Delta\hat{\mathcal{Y}}_d/\tau)}{\sum_{h=1}^{H}\sum_{w=1}^{W}\exp(\Delta\hat{\mathcal{Y}}_{d,h,w}/\tau)}, \tag{11}$$

$$\sigma(\Delta\mathcal{Y}_d) = \frac{\exp(\Delta\mathcal{Y}_d/\tau)}{\sum_{h=1}^{H}\sum_{w=1}^{W}\exp(\Delta\mathcal{Y}_{d,h,w}/\tau)}, \tag{12}$$

  - $\mathcal{L}_{\text{Div}}$: Kullback-Leibler divergence between difference maps:

$$\mathcal{L}_{\text{Div}} = \sum_{d=1}^{D-1}\sigma(\Delta\hat{\mathcal{Y}}_d)\log\frac{\sigma(\Delta\hat{\mathcal{Y}}_d)}{\sigma(\Delta\mathcal{Y}_d)} \tag{13}$$

Overall Flow



Experiment

THE CHIPS TO SYSTEMS CONFERENCE

Experimental Setup

- Mask clip: $2 \times 2\mu m^2$ with $80nm$ thickness
- Resolution (x, y, z): $2nm$, $2nm$, $1nm$.
- Technology node: $28nm$ and below
- Simulation parameter:

Table: Important parameters in photoresist simulation process.

PEB			
Normal Diffusion Length $L_{N,\mathcal{A}}, L_{N,\mathcal{B}}$	70, 15 nm	Lateral Diffusion Length $L_{L,\mathcal{A}}, L_{L,\mathcal{B}}$	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 $[\mathcal{A}]_{sat}, [\mathcal{B}]_{sat}$	0.9, 0
$[\mathcal{I}](t=0)$	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		

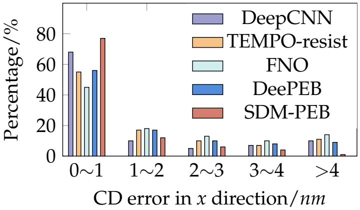
Compare With Learning-based PEB Solvers

- DeepCNN: convolutional neural network model with a residual connection
- TEMPO-resist: conditional-GAN based model
- FNO: Fourier neural network
- DeePEB: extends FNO with CNN-based local learning branches

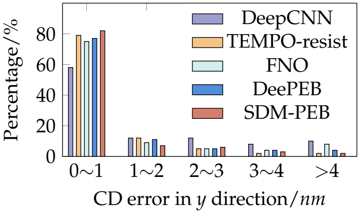
Table: Comparison with different PEB solvers.

Methodologies	Inhibitor RMSE (*e*-3)	Inhibitor NRMSE (%)	Develop Rate RMSE (*nm*/*s*)	Develop Rate NRMSE (%)	CD Error x (*nm*)	CD Error y (*nm*)	RT/s
DeepCNN	8.25	12.53	0.65	1.63	3.14	6.26	**1.01**
TEMPO-resist	7.67	12.55	0.50	1.26	2.12	2.45	6.48
FNO	7.91	11.68	0.68	1.69	2.34	3.71	1.15
DeePEB	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

Comparison of CD Error



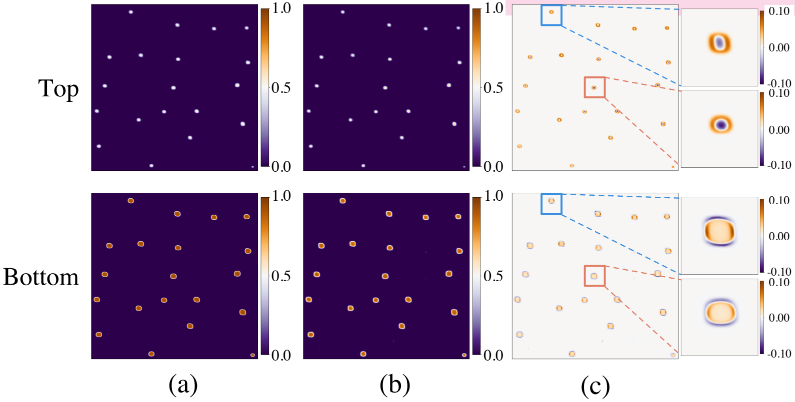
(a)



(b)

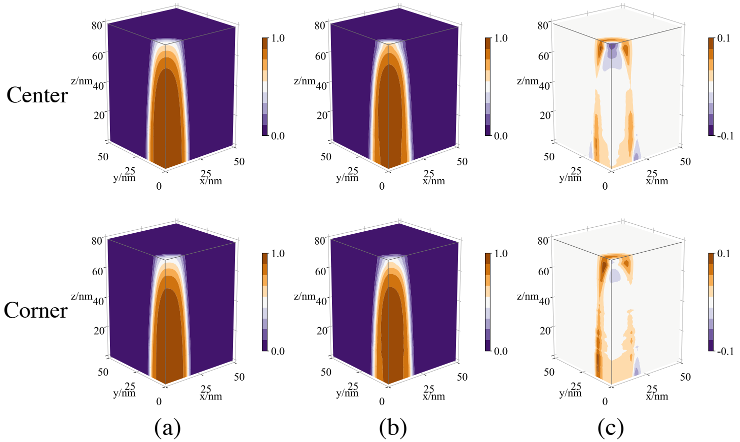
Percentage counts of CD errors using different methods: (a) error in the x direction and (b) error in the y direction.

Visualization of Simulation Results



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.

Visualization of Simulation Results



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.

Ablation Study

Table: Ablation study

Methodologies	NRMSE/%		CD Error	
	Inhibitor	Rate	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**

62

THE CHIPS TO SYSTEMS CONFERENCE

SPONSORED BY CEDA sigda