# 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


## Highlights

- **Precise & Fast PEB Modeling:** We introduce **SDM-PEB**, a streamlined pipeline for accurate, high-speed post-exposure-bake simulation.
- **Hierarchical ViT Encoder:** Extracts multi-scale spatial features within each photoacid depth layer.
- **Spatial-Depthwise Mamba Attention:** A lightweight unit that captures cross-depth dependencies efficiently.
- **PEB Focal Loss & Depthwise Divergence Reg.:** Tailored objectives that handle data imbalance and refine layer-wise learning.
- **Industry Validation:** On *S-Litho* benchmarks, SDM-PEB outperforms existing methods in both accuracy and runtime.

## Background and Motivaton

**Predictive Lithography Simulation** comprises two key stages:

- Optical Simulation: models ligh-mask interaction and pattern projection onto the photoresist.
- Photoresist Simulation: captures resist chemistry and physics from exposure, through post-exposure bake (PEB), to final development.

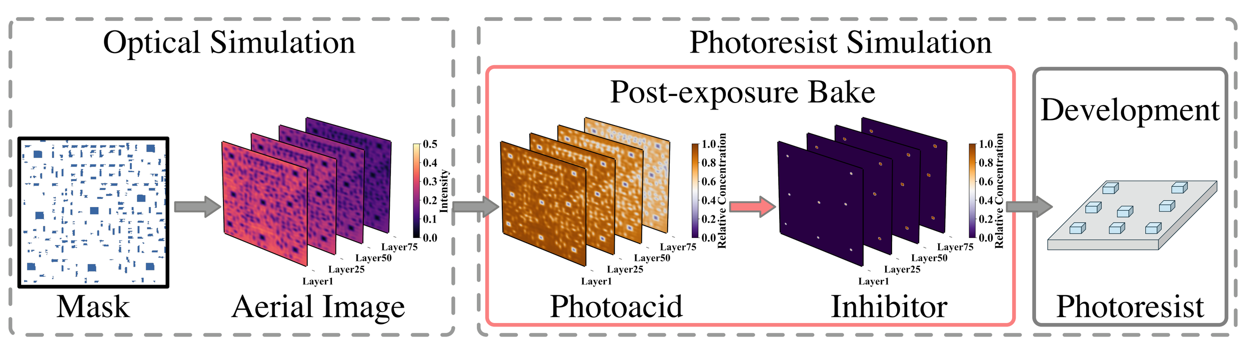


Figure 1. A typical flow of lithography simulation for chemically amplified resist: from optical simulation to photoresist simulation.

**Post-Exposure Bake Process:** Thermal-driven PEB process for chemically amplified resist contain

1. Catalytic reaction between inhibitor $[\mathcal{I}]$ and photoacid $[\mathcal{A}]$ Equation (1),
2. neutralization-diffusion between photoacid and base quencher $[\mathcal{B}]$ Equation (2),
3. photoresist development with rate $R$ Equation (4).

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

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

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

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

**Existing Problems:**

- Finite element analysis (FEA) and finite difference methods (FDM) demand significant computational resources due to 3D distribution characteristic.
- DeePEB [2] cannot fully capture continuous spatial and depthwise dependencies in 3D space using FNO and CNN.

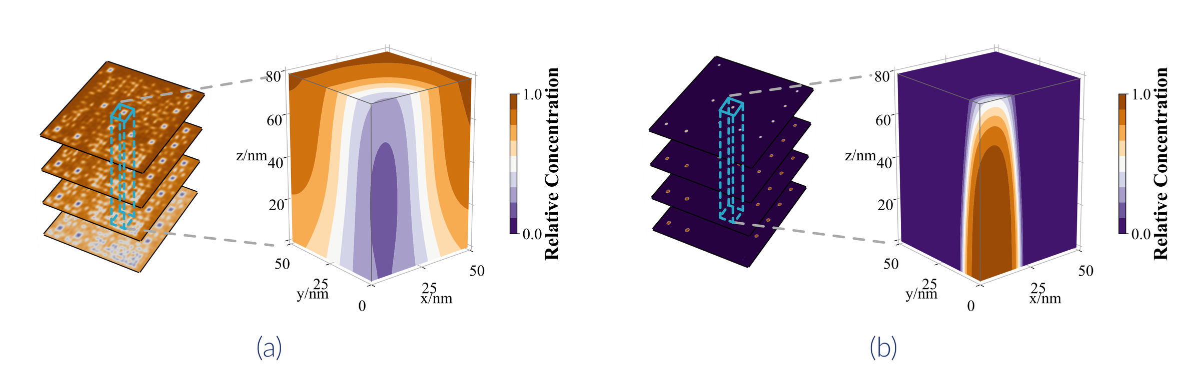


Figure 2. Vertical visualization of distributions: (a) photoacid at the initial stage and (b) inhibitor at the final stage.

## State Space Models-based Methodologies

- Feature representations in 3D photoresist can be naturally modeled as sequences of depth-levels (from shallow to deep).
- **Linear time-invariant SSM:** maps scalar input sequence $x_t$ to scalar output $y_t$ through hidden state $h_t \in \mathbb{R}^N$, for $t = 1, \ldots, L$.
- **Dynamics:** $h_{t+1} = A\,h_t + B\,x_t, \quad y_t = C^\top h_t.$
- **Parameters:**
  - $A \in \mathbb{R}^{N\times N}$ — state matrix, HiPPO-initialized.
  - $B, C \in \mathbb{R}^{N\times 1}$ — input/output projection vectors.
- **Mamba** ties SSM parameters to the input, spotlighting relevant signals and filtering out the rest.

$$\vec{B} = \texttt{Linear}_N(\vec{x}),\ \vec{C} = \texttt{Linear}_N(\vec{x}), \tag{5}$$

$$\vec{\Delta} = \texttt{softplus}(\texttt{Broadcast}_K(\texttt{Linear}_1(\vec{x})) + \vec{D}), \tag{6}$$

## Overall flow

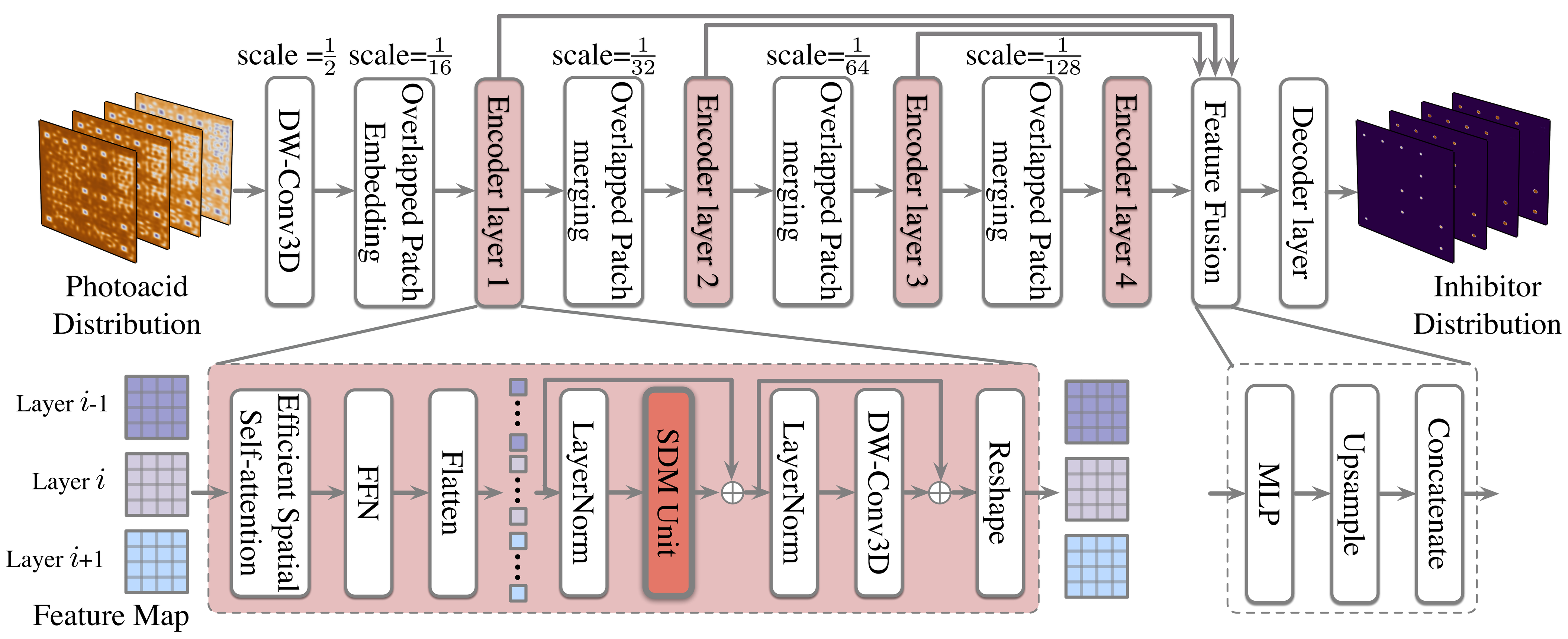


Figure 3. The architecture overview of our proposed SDM-PEB framework.

## Hierarchical Contextual Feature Extractor

**Depthwise Overlapped Patch Merging:**

- Reduce information loss at patch boundaries

**Efficient Spatial Self-Attention:**

- $C$: feature dimension of $\vec{K}$; $r$: reduction ratio
- Computational complexity: $O(L^2) \to O(L^2/r)$

$$\hat{\vec{K}} = \texttt{Reshape}\left(\frac{L}{r}, C\cdot r\right)(\vec{K}), \quad \vec{K} = \texttt{Linear}_C(\hat{\vec{K}}), \tag{7}$$

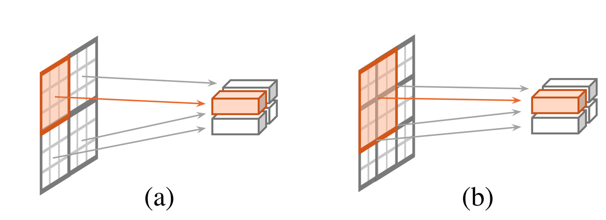


Figure 4. (a) Non-overlapped patch merging and (b) overlapped patch merging.

## Spatial-Depthwise Mamba-based Attention Unit

**Structure of SDM Unit:**

- **Input reshaping:** The $i$-th layer feature map is flattened to a sequence $q_i$ and normalized.
- **Dual projection:** $q_i$ is linearly split into two streams $x_i$ and $z_i$.
- **Directional scans:** For every scan direction, $x_i$ passes through a 1-D conv + SiLU to yield $x_i'$, which drives a spatial-depthwise selective scan.
- **Attention fusion:** Scan outputs are gated by $z_i$ and aggregated to form feature $p_i$, capturing inter-/intra-layer dependencies.

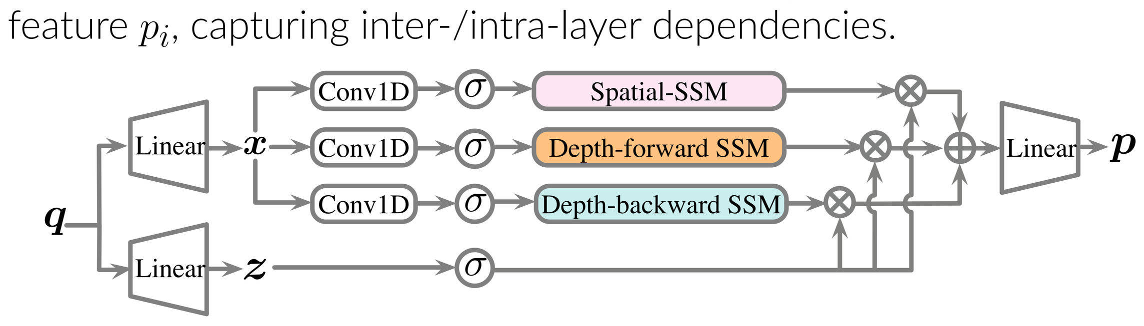


Figure 5. The architecture of the spatial-depthwise Mamba-based attention unit.

**Spatial-Depthwise PEB Selective Scan:**

- Spatial scan: at a fixed $(x, y)$ location, sweeps through all depth layers.
- Depth-Forward Scan: traverses layers shallow → deep
- Depth-Backward Scan: traverses layers deep → shallow (reverse of depth-forward)

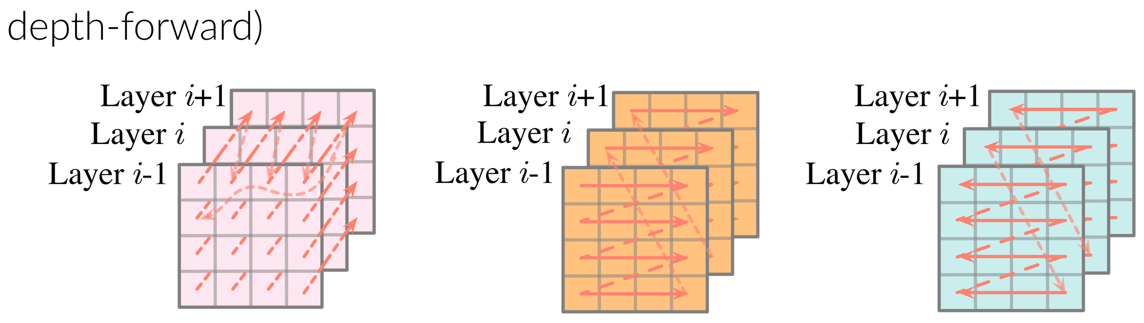


Figure 6. From left to right: spatial scan, depth-forward scan, and depthbackward scan.

## Customized PEB Optimization Objectives

- **Maximum squared error:** $\mathcal{L}_{\text{MaxSE}} = \max_{d,h,w}\left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right)^2.$
- **PEB focal loss:**
  $\mathcal{L}_{\text{PEB-FL}} = \sum_d^D \sum_h^H \sum_w^W \left|\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right|^\gamma \left(\hat{\mathcal{Y}}_{d,h,w} - \mathcal{Y}_{d,h,w}\right)^2$
- **Differential Depth Divergence Regularization:**
  $\mathcal{L}_{\text{Div}} = \sum_{d=1}^{D-1} \sigma(\Delta\hat{\mathcal{Y}}_d) \log\frac{\sigma(\Delta\hat{\mathcal{Y}}_d)}{\sigma(\Delta\mathcal{Y}_d)}$, where:
  $\Delta\mathcal{Y}_d = \mathcal{Y}_{d+1} - \mathcal{Y}_d, \ \ \sigma(\Delta\mathcal{Y}_d) = \frac{\exp(\Delta\mathcal{Y}_d/\tau)}{\sum_{h=1}^H\sum_{w=1}^W \exp(\Delta\mathcal{Y}_{d,h,w}/\tau)}$

## Evaluation Results

Table 1. 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 [3] | 8.25 | 12.53 | 0.65 | 1.63 | 3.14 | 6.26 | **1.01** |
| TEMPO-resist [4] | 7.67 | 12.55 | 0.50 | 1.26 | 2.12 | 2.45 | 6.48 |
| FNO [1] | 7.91 | 11.68 | 0.68 | 1.69 | 2.34 | 3.71 | 1.15 |
| DeePEB [2] | 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 |

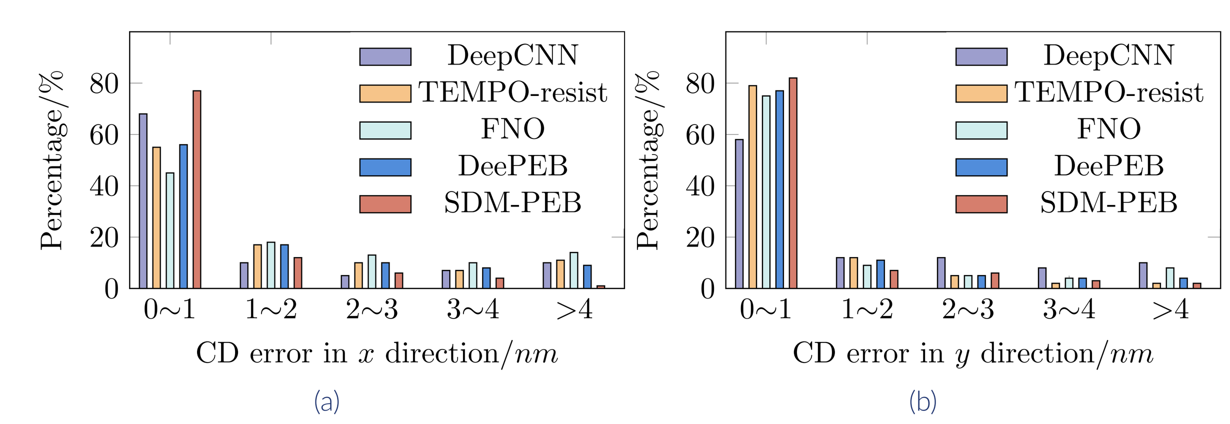


Figure 7. Percentage counts of CD errors using different methods: (a) error in the $x$ direction and (b) error in the $y$ direction.

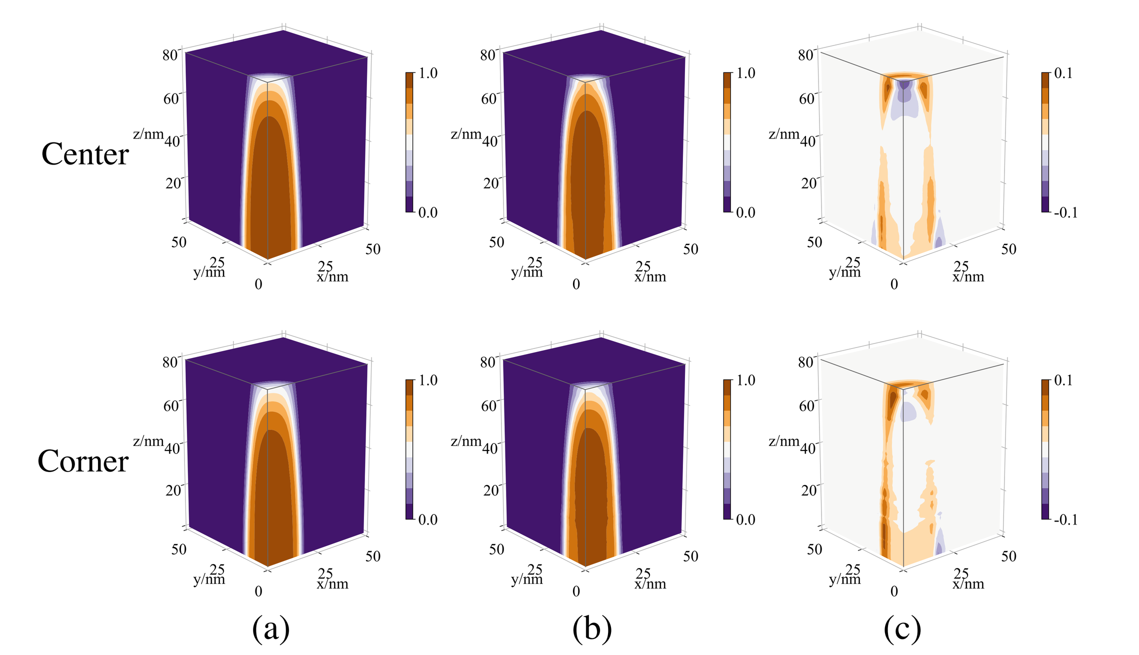


Figure 8. Vertical visualization of predicted results: the upper row shows the center contact, the lower row shows the corner contact. (a) Ground truths, (b) predictions, (c) differences.

## References


[1] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[2] Qipan Wang, Xiaohan Gao, Yibo Lin, Runsheng Wang, and Ru 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*, pages 1–9, 2022.

[3] Yuki Watanabe, Taiki Kimura, Tetsuaki Matsunawa, and Shigeki Nojima. Accurate lithography simulation model based on convolutional neural networks. In *Optical Microlithography XXX*, volume 10147, pages 137–145. SPIE, 2017.

[4] Wei Ye, Mohamed Baker Alawieh, Yuki Watanabe, Shigeki Nojima, Yibo Lin, and David Z Pan. Tempo: Fast mask topography effect modeling with deep learning. In *Proceedings of the 2020 International Symposium on Physical Design*, pages 127–134, 2020.