DiffResist: Physics-Constrained Diffusion for Photoresist Modeling

Zixiao Wang[1], Jieya Zhou[1], Xinyun Zhang[1], Shoubo Hu[2], Farzan Farnia[1], Bei Yu[1]
[1]CUHK [2]Huawei

***Abstract*—Accurate and efficient 3D photoresist simulation is essential for optical lithography at advanced technology nodes. Existing methods that predict 3D resist profiles from aerial images either rely on analytical reaction–diffusion solvers, which are slow, or on high-capacity 3D generative models, which are costly to train and deploy. We instead formulate 3D resist prediction as a depth-wise 2D generation task conditioned on the aerial image.** DiffResist **introduces a physics-constrained diffusion model whose reverse steps are aligned with resist exposure physics: a two-stage noise schedule connects physically meaningful layers to a Gaussian prior, and boundary conditions at the resist–air interface are injected to suppress error propagation. Combined with a lightweight super-resolution module,** DiffResist **achieves state-of-the-art accuracy on a public benchmark with over 10× faster inference than 3D diffusion baselines.**

I. Introduction

Optical lithography dominates both the cost and performance of modern integrated-circuit manufacturing, accounting for a large fraction of total fabrication cost [1]–[5]. Accurate, fast simulation is therefore indispensable for yield control and turnaround time reduction, and the photoresist process is especially critical: imperfect resist modeling directly translates into edge-placement errors at advanced nodes.

Resist simulation is commonly divided into 2D and 3D modeling. 2D approaches predict top-down or cross-sectional profiles from pre-computed aerial images using threshold-based or data-driven models [3], [6]–[10]. They are lightweight but cannot capture complex photo-chemical phenomena in thick resists and EUV settings. 3D models simulate volumetric resist profiles via coupled reaction–diffusion equations and related analytical formulations [11]–[14]. While physically interpretable, such solvers are slow to evaluate and increasingly difficult to calibrate at nanoscale dimensions.

Learning-based methods offer an attractive alternative. The task can be viewed as predicting a sequence of 2D inhibitor concentration maps along depth, conditioned on a given aerial image, reminiscent of video prediction [15]–[18]. Direct 3D diffusion models are computationally heavy, while purely sequential predictors (e.g., recurrent or autoregressive models) suffer from error accumulation along depth [19], [20]. Unlike open-ended video generation, 3D resist simulation benefits from well-studied exposure physics that can act as strong priors to regularize learning and control error propagation.

To exploit these physical priors while maintaining efficiency, we propose DiffResist, a physics-constrained diffusion framework for 3D photoresist profile simulation. Instead of learning a full 3D mapping, DiffResist performs diffusion on a down-sampled 2D grid and aligns each reverse step with one physical layer in the resist stack. A two-stage noise schedule separates physically meaningful layers from auxiliary warm-up steps that connect the bottom layer to a Gaussian prior. Boundary conditions at the resist–air interface are injected into the reverse process to correct the top layer, and a lightweight residual super-resolution network upsamples the low-resolution stack to the target resolution.

In short, DiffResist (i) reformulates volumetric prediction as a physics-aligned 2D diffusion process, (ii) integrates resist exposure physics and a two-stage noise schedule to mitigate error accumulation along depth, and (iii) uses a small super-resolution head to reconstruct high-resolution 3D profiles, achieving state-of-the-art accuracy with over $10\times$ faster inference than 3D diffusion baselines.

II. DiffResist: Physics-Constrained Diffusion

A. Overview and Physics

We consider a standard positive-tone resist exposure model [21], where light intensity $I(h,t)$ and normalized inhibitor concentration $M(h,t) \in [0,1]$ satisfy coupled reaction–diffusion equations along depth h and time t [11], [12]. The aerial image $R(x,y)$ acts as a boundary condition at the resist surface $h=0$. In particular, for a given exposure duration T, the top-layer inhibitor at the air–resist interface has the closed form $M_0(R) = \exp(-RCT)$ for a known material constant C. Classical simulators step these equations in time and then apply a bulk development model [22] to derive the final resist profile, which is accurate but time-consuming.

DiffResist instead directly predicts the final fractional inhibitor volume $M(h,T)$ along depth, discretized as slices $M_0,\dots,M_{D-1}$ from the resist surface to the bottom. As shown in Fig. 1, the method has two stages: a physics-constrained 2D diffusion model that predicts a low-resolution 3D inhibitor stack, and a lightweight super-resolution network that reconstructs the final high-resolution volume.

B. Depth-Wise 2D Diffusion

We represent the 3D inhibitor volume as a stack of D 2D slices $M_d \in [0,1]^{H\times W}$ and run diffusion on a down-sampled grid. During training, slices $\{M_d^{\text{data}}\}_{d=0}^{D-1}$ are available from a simulator. The forward process is defined so that physical layers $M_{0:D-1}$ remain deterministic (equal to simulator outputs), while an extra short segment of length D_E is diffused towards

This work is partially supported by Research Grants Council of Hong Kong SAR (No. RFS2425-4S02).

a Gaussian prior, following the DDPM construction [23]. The last state in this segment approximates $\mathcal{N}(0, I)$.

The reverse process starts from this Gaussian state and sequentially predicts M_{d-1} from M_d conditioned on the aerial image R. We use a noise-prediction network $\boldsymbol{\epsilon}_\theta(M_d, d, R)$ and a closed-form Gaussian update as in DDPM, but with one key change: the top layer is always anchored to the analytic boundary condition instead of being fully learned. The reverse update at step d is

$$M_{d-1} = \sqrt{\bar{\alpha}_{d-1}}\, M_0(R) + \kappa_d\, \boldsymbol{\epsilon}_\theta(M_d, d, R) + \sigma_d z, \qquad (1)$$

where $\bar{\alpha}_{d-1}$ and κ_d are pre-computed scalars, $z \sim \mathcal{N}(0, I)$, the first term injects the exact surface profile $M_0(R)$, and the second term provides a learned directional correction.

To reflect that $M_{0:D-1}$ are deterministic given R, we use a two-stage variance schedule: for steps corresponding to physical layers we set $\sigma_d^2 \approx 0$ (removing sampling noise and preventing error accumulation along depth), whereas for the auxiliary warm-up segment we gradually increase σ_d^2 to connect the bottom layer to the Gaussian prior. Training follows the standard DDPM noise-prediction objective [23]. At inference, we start from a Gaussian sample, apply Equation (1) for all steps, and finally take the first D slices as the predicted low-resolution 3D inhibitor volume.

C. Lightweight Super-Resolution

To further reduce cost, DiffResist operates on spatially down-sampled inputs and predicts a low-resolution stack $M_{\text{low}} \in [0,1]^{H' \times W' \times D}$. The resist inhibitor field is smoother than natural images, so high-resolution reconstruction from M_{low} is feasible. We employ a lightweight residual super-resolution network of the form

$$M_{\text{high}} = g(M_{\text{low}}, R) + \text{UP}(M_{\text{low}}),$$

where g is a small CNN conditioned on the aerial image R, and UP denotes non-parameterized bilinear interpolation. Down-sampling by factors $(\lambda_H, \lambda_W, \lambda_D)$ along height, width, and depth yields an approximate $\lambda_H \lambda_W \lambda_D$ reduction in compute. After super-resolution, the final printed pattern is obtained via a bulk development model [22], using M_{high} as input.

III. Experiments and Conclusion

We evaluate DiffResist on LithoBench [24], which contains 16,472 tiles with ground-truth 3D resist profiles from the analytical simulator TorchResist [25]. We use its 2000-step outputs as references. The simulator produces inhibitor volumes at 7 nm/pixel with size $(75, 147, 147)$; in DiffResist, we down-sample by $(\lambda_H, \lambda_W, \lambda_D) = (3, 3, 5)$ to $(15, 49, 49)$, set $D = 15$ physical layers and a short warm-up segment, and adopt a U-Net backbone [26] with a two-stage noise schedule. A small dual-branch residual super-resolution network is conditioned on the aerial image.

We compare against two direct 3D diffusion baselines, 3D-DDPM [23] and 3D-DDIM [27], and two sequential baselines, SVG [16] and SVG-RNN [28]. All methods operate on the same down-sampled volume and use bilinear upsampling when needed. We report mean squared error (MSE) between predicted and reference inhibitor volumes, Pixel Difference (PD) between binarized printed images [25] in percent, and edge placement error (EPE-mean and EPE-max, in nm) [6].

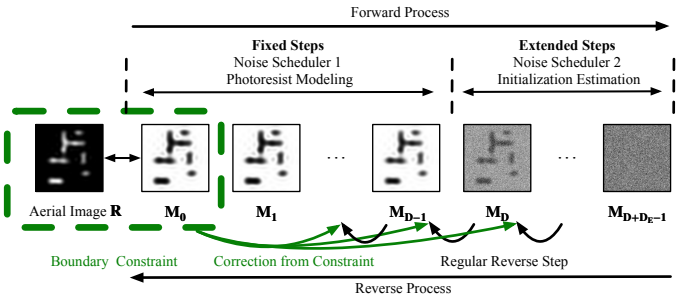

Fig. 1 Overview of DiffResist. A physics-constrained 2D diffusion model predicts a low-resolution 3D inhibitor volume, which is then upsampled by a super-resolution network.

TABLE I Results on LithoBench [24]. PD: Pixel Difference; EPE in nm.

Method	MSE	PD	EPE-max	EPE-mean
3D-DDPM	4.14×10^{-2}	19.10%	40.47	10.17
3D-DDIM	3.79×10^{-2}	18.54%	39.52	9.86
SVG	1.41×10^{-3}	2.18%	19.14	7.73
SVG-RNN	7.39×10^{-4}	1.02%	17.06	4.62
DiffResist	$\mathbf{4.17 \times 10^{-4}}$	**0.85%**	**13.64**	**2.56**

TABLE I summarizes the quantitative results. Direct 3D diffusion methods (3D-DDPM and 3D-DDIM) incur relatively large errors (MSE $\approx 4 \times 10^{-2}$, PD $\approx 19\%$, EPE-mean ≈ 10 nm) despite using large 3D models. Sequential baselines (SVG and SVG-RNN) are more efficient and substantially more accurate, but their purely autoregressive design is still affected by error accumulation along depth.

In contrast, DiffResist achieves the best accuracy on all metrics while keeping the model compact. It reduces MSE to 4.17×10^{-4} and PD to 0.85%, and lowers EPE-mean from 4.62 nm (SVG-RNN) to 2.56 nm and EPE-max from 17.06 nm to 13.64 nm. In terms of runtime, 3D-DDPM needs 6.45 s per tile with EPE-mean 10.17 nm, whereas DiffResist attains EPE-mean 2.56 nm in about 0.44 s per tile, i.e., over $10\times$ faster for markedly better accuracy.

DiffResist reforms 3D photoresist profile prediction as a physics-constrained 2D diffusion problem. By aligning diffusion steps with physical depth, injecting the resist–air boundary condition into the reverse process, and confining stochasticity to a short warm-up segment, it mitigates error accumulation that plagues generic sequential models. A lightweight residual super-resolution network further reduces computational cost by operating on down-sampled volumes. Experiments on LithoBench with TorchResist references show that DiffResist achieves state-of-the-art accuracy and over $10\times$ faster inference than 3D diffusion baselines; extending the framework with additional physical priors and to transient or stochastic resist processes is a promising direction for future work.

REFERENCES

[1] C. Mack, *Fundamental principles of optical lithography: the science of microfabrication*. John Wiley & Sons, 2007.

[2] Z. Wang, Y. Shen, W. Zhao, Y. Bai, G. Chen, F. Farnia, and B. Yu, "Diff-pattern: Layout pattern generation via discrete diffusion," in *ACM/IEEE Design Automation Conference (DAC)*, 2023.

[3] 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.

[4] Z. Wang, Y. Shen, X. Yao, W. Zhao, Y. Bai, F. Farnia, and B. Yu, "ChatPattern: Layout Pattern Customization via Natural Language," in *ACM/IEEE Design Automation Conference (DAC)*, 2024, pp. 1–6.

[5] 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 (TCAD)*, 2024.

[6] S. Banerjee, Z. Li, and S. R. Nassif, "ICCAD-2013 CAD contest in mask optimization and benchmark suite," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2013, pp. 271–274.

[7] J. Randall, K. G. Ronse, T. Marschner, A.-M. Goethals, and M. Ercken, "Variable-threshold resist models for lithography simulation," in *Optical Microlithography XII*, vol. 3679. SPIE, 1999, pp. 176–182.

[8] C. A. Mack, "A new fast resist model: the gaussian lpm," in *Design for Manufacturability through Design-Process Integration V*, vol. 7974. SPIE, 2011, pp. 65–70.

[9] Y. Granik, N. B. Cobb, and T. Do, "Universal process modeling with vtre for opc," in *Optical Microlithography XV*, vol. 4691. SPIE, 2002, pp. 377–394.

[10] F. X. Zach, "Neural-network-based approach to resist modeling and opc," in *Optical Microlithography XVII*, vol. 5377. SPIE, 2004, pp. 670–679.

[11] J. Thackeray, J. Cameron, V. Jain, P. LaBeaume, S. Coley, O. Ongayi, M. Wagner, A. Rachford, and J. Biafore, "Pursuit of lower critical dimensional uniformity in euv resists," *Journal of Photopolymer Science and Technology*, vol. 26, no. 5, pp. 605–610, 2013.

[12] M. D. Smith, J. D. Byers, and C. A. Mack, "The lithographic impact of resist model parameters," in *Advances in Resist Technology and Processing XXI*, vol. 5376. SPIE, 2004, pp. 322–332.

[13] J. W. Thackeray, R. A. Nassar, R. Brainard, D. Goldfarb, T. Wallow, Y. Wei, J. Mackey, P. Naulleau, B. Pierson, and H. H. Solak, "Chemically amplified resists resolving 25 nm 1: 1 line: space features with euv lithography," in *Emerging Lithographic Technologies XI*, vol. 6517. SPIE, 2007, pp. 394–404.

[14] J. W. Thackeray, "Materials challenges for sub-20-nm lithography," *Journal of Micro/Nanolithography, MEMS, and MOEMS (JM3)*, vol. 10, no. 3, pp. 033 009–033 009, 2011.

[15] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "Simvp: Simpler yet better video prediction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 3170–3180.

[16] E. Denton and R. Fergus, "Stochastic video generation with a learned prior," in *International Conference on Machine Learning (ICML)*. PMLR, 2018, pp. 1174–1183.

[17] Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, "Eidetic 3D LSTM: A model for video prediction and beyond," in *International Conference on Learning Representations (ICLR)*, 2018.

[18] S. Chen, M. Xu, J. Ren, Y. Cong, S. He, Y. Xie, A. Sinha, P. Luo, T. Xiang, and J.-M. Perez-Rua, "Gentron: Diffusion transformers for image and video generation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 6441–6451.

[19] L. Hu, "Animate anyone: Consistent and controllable image-to-video synthesis for character animation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 8153–8163.

[20] G. Shrivastava and A. Shrivastava, "Video prediction by modeling videos as continuous multi-dimensional processes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 7236–7245.

[21] 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.

[22] C. A. Mack, "Development of positive photoresists," *Journal of the Electrochemical Society*, vol. 134, no. 1, p. 148, 1987.

[23] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Annual Conference on Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[24] S. Zheng, H. Yang, B. Zhu, B. Yu, and M. Wong, "LithoBench: Benchmarking ai computational lithography for semiconductor manufacturing," vol. 36, 2024.

[25] Z. Wang, J. Zhou, S. Zheng, S. Yin, K. Liang, S. Hu, X. Chen, and B. Yu, "Torchresist: Open-source differentiable resist simulator," in *SPIE Advanced Lithography + Patterning*, 2025.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2015, pp. 234–241.

[27] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[28] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.