A Diffusion Model enhanced Source Optimization for Fab Productivity

Xiaoxiao Liang[a], Benjamin C. He[b], and Yuzhe Ma[a*]

[a]The Hong Kong University of Science and Technology (Guangzhou), China
[b]University of California, Berkeley, United States

ABSTRACT

In the photolithography process at advanced nodes, source optimization (SO) becomes crucial for enhancing the production yield of each mask. Due to the complexity of physical models as well as the multi-criteria global optimization per full-chip mask, the overall SO procedure for each mask is both time-consuming and resource-intensive in production. To enhance the fab efficiency and productivity, we propose a new, diffusion-based SO procedure. The framework is established on a latent diffusion model focusing on text-to-image generation. We generated hundreds of layout-source pairs using an in-house physical SO tool as the training dataset, targeting 28nm nodes. The process parameters and the layout characteristic data are organized as textual inputs to the diffusion model. They are then encoded for conditional image generation, with source images serving as the optimization reference. A complete training-and-inference pipeline is presented, along with visualization of testcases and additional observations.

Keywords: Source optimization, latent diffusion model

1. INTRODUCTION

Optical projection lithography remains the primary method for defining critical features on semiconductor wafers. As device dimensions shrink below the exposure wavelength, maintaining imaging fidelity, depth of focus (DOF), and exposure latitude (EL) becomes increasingly difficult due to diffraction limits. Traditional resolution enhancement techniques (RETs) such as Optical Proximity Correction (OPC) and Phase-Shift Masks (PSM), together with off-axis illumination (OAI), high-contrast photoresists, and advanced patterning processes, have enabled scaling to 14nm nodes and beyond. However, to further approach theoretical resolution limits, holistic co-optimization of all process variables—particularly the illumination source and photomask—is essential. Computational lithography methods, notably Source Mask Optimization (SMO) and Inverse Lithography Technology (ILT), which leverage detailed lithography simulations for end-to-end process optimization, entered industrial use around 2009 and have become critical for meeting future node requirements.

SMO jointly optimizes the illumination source and photomask patterns. It is designed as an essential method to enhance lithographic performance by controlling imaging fidelity, depth of focus (DOF), exposure latitude (EL), and mask enhancement factor (MEEF).[1,2] Industry-standard approaches emphasize integrating programmable illumination systems—such as off-axis dipole or quadrupole sources—with optimized mask features, leading to demonstrated DOF improvements of up to 50% and exposure latitude gains of 30% in practical DUV processes.[3,4] Theoretical advancements in SMO have significantly enhanced lithographic robustness and precision.

Jia and Lam[5] introduced a pixelated source-mask framework, achieving a 25% reduction in critical dimension (CD) variation across process windows by modulating 64×64 source pixels. Ma et al.[6] incorporated vectorial imaging models to account for polarization-dependent diffraction, reporting a 15% improvement in aerial image contrast at k1 values below 0.35. Further, hybrid optimization methods combining L2-norm minimization with augmented Lagrangian techniques[7] reduced mask feature bias to < 2 nm while maintaining sub-1% computational overhead increase compared to scalar models. Gradient-based source tuning approaches[8] achieved convergence within 50 iterations, illustrating efficiency gains for full-chip SMO tasks. Level set methods, known

* Corresponding author, e-mail: yuzhema@hkust-gz.edu.cn

for their flexibility in handling complex geometrical evolution problems, have been effectively adapted for inverse lithography techniques (ILT) and mask optimization. Zou[9] reported that ILT approaches using level set formulations enhanced CD uniformity by 20% and improved sidewall angle fidelity by 10% at 22 nm nodes. Shen et al.[10,11] applied robust level-set algorithms with regularization terms to manage mask topology changes, achieving < 5 nm edge placement error (EPE) under stochastic process variations. These methods demonstrated mask manufacturability improvements by reducing sub-resolution assist feature (SRAF) count by 30% while preserving target critical dimensions.

Practical applications of SMO require comprehensive workflows integrating optimization tools and methodologies. Rosenbluth et al.[12] demonstrated a holistic SMO flow combining source shaping, mask biasing, and illumination optimization, yielding a 40% DOF increase for 22 nm node patterns. Xiao[13] focused on minimizing mask error enhancement factor (MEEF), achieving a reduction from 2.5 to 1.8 across multiple pattern types. Semiconductor fabs leveraging these workflows reported up to a $2\times$ improvement in lithographic yield. Lafferty et al.[14,15] evaluated RET selection in NAND flash processes, highlighting that integrating SMO with OPC reduced systematic CD error by 22%. Zhang et al.[16] applied full-chip SMO and OPC co-optimization, observing a 15% reduction in line-edge roughness (LER) and a 25% improvement in yield.

All of the physical model–based SO/SMO methods described above demand substantial computational resources and setup time — often taking days or even weeks to complete a full optimization cycle. In contrast, recent advances in machine learning (ML) have dramatically accelerated SMO workflows while maintaining or improving precision, effectively addressing both runtime overhead and complexity challenges. Chen et al.[17] developed physics-informed optical kernel regression with complex-valued neural fields, reducing model evaluation time by 60% and achieving < 3 nm mean CD error compared to rigorous simulations. Yang et al.[18] used Convolutional Fourier Neural Operators (CFNO) with lithography-guided self-training, enabling full-chip mask optimization in under 10 minutes—over $10\times$ faster than iterative methods—while maintaining < 2 nm CD error. DevelSet[19] employs a deep neural level set that converges in < 1 s per feature, achieving < 1.5 nm EPE. DAMO[20] integrates deep lithography simulations and mask generation, demonstrating a $5\times$ speedup in full-chip optimization and a 20% improvement in process window size. Although ML-driven computational lithography is still in an exploratory phase—working to validate its practicality in chip manufacturing—its ability to boost precision and reduce runtimes has captured the attention of leading wafer fabs.

To address day-to-day production challenges faced by lithography process engineers, this research emphasizes source-only optimization within an overall manufacturable reticle. Reducing warm or hot spots and improving yield is a fundamental job requirement in production fabs. Leveraging existing SMO technologies but focusing solely on source optimization has demonstrated significant lithographic improvements in many case studies. Wu et al.[21] introduced freeform source profiles tailored to 'warm spots,' resulting in a 35% reduction in localized CD hotspots and a 25% increase in uniformity across dense pattern arrays. Yu et al.[22] combined margin image averaging with conjugate gradient optimization, achieving a 30% gain in exposure latitude and a 0.8nm reduction in CD mean bias. These targeted source-optimization approaches enable rapid turnaround and effectively complement full SMO workflows.

In this paper, we propose a novel diffusion-based framework for source optimization (SO) in lithography. Leveraging a latent diffusion model for text-to-image–style source pattern generation, our approach provides lithography engineers with an intuitive and efficient tool for mitigating hot spots and improving yield. The remainder of this paper is organized as follows: (1) we review the benefits of off-axis illumination (OAI) and describe the generation of training data using OAI-based optical simulations across a range of pattern densities; (2) we present the architecture and training process of the latent diffusion network, incorporating lithographic constraints; (3) we outline the end-to-end training and inference pipeline, including visualization tools for rapid deployment; and (4) we provide case studies demonstrating the practicality, usability, and effectiveness of the proposed diffusion-model based, SO framework for fab production teams in daily operations.

2. PRELIMINARIES

In optical lithography, resolving increasingly smaller features at advanced technology nodes demands precise control of not only the projection optics but also the illumination source. The illumination configuration—defined

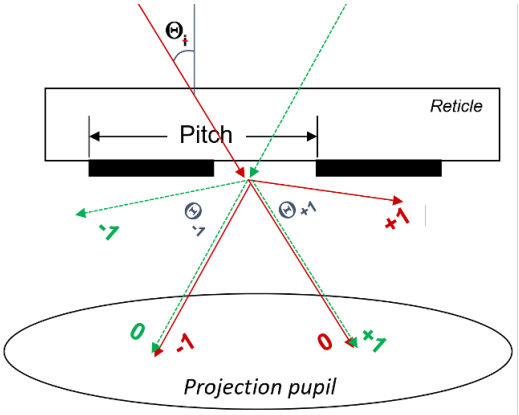


Figure 1: 0th and 1st diffraction orders are spread symmetrically inside the lens NA, due to the optimized sigma (σ) value of an OAI.

by its angular and spatial distribution in the pupil plane—plays a pivotal role in determining how diffraction orders from mask features propagate through the lens system and contribute to image formation on the wafer. Among the various RET methods, OAI has become essential for enabling two-beam interference imaging that captures critical diffraction orders with high contrast and extended DOF.[23]

Depth of focus refers to the range over which the imaging system can maintain an acceptable focus on the photoresist layer, which is a critical variable during a lithography process setup. In an optical system, allowing multiple diffraction orders to pass through the lens generally improves imaging quality by enhancing resolution and contrast. However, this comes at the cost of reduced depth of focus, as the overlapping focus range of all the contributing diffraction orders becomes narrower. A key benefit of OAI is that it can be configured to allow only the 0th and $\pm$1st diffraction orders of a given pitch to enter the lens, symmetrically, forming a two-beam interference pattern on the wafer. OAI achieves this by shifting the illumination source off the optical axis, often into discrete poles positioned at carefully selected angles. This approach improves the optical transfer function (OTF) at specific spatial frequencies, especially for periodic grating structures common in logic and memory layouts. Two-beam imaging enhances image contrast (by scarifying partial image fidelity) and extends the DOF far beyond what would be possible with traditional on-axis or broad-field illumination.

Dipole illumination is a fundamental type of OAI where two coherent source points are placed symmetrically on opposite sides of the optical axis. This setup is particularly effective for printing line-space patterns and other 1D-dominant features. The theoretical framework suggests that two-beam imaging using perfect dipoles can achieve arbitrarily long DOF as long as the beams are perfectly coherent and symmetric. In practice, realizing this benefit requires precise control of the sigma (σ) value, which determines the angular displacement of the illumination poles relative to the optical axis. Correctly choosing σ ensures that the 0th and 1st diffraction orders are spread symmetrically inside the lens NA, maximizing DOF, therefore the process stability (Fig. 1).

For such a dipole-based OAI targeting two-beam imaging, the optimal sigma value σ is in fact given by a simple formula:

$$\sigma = \frac{\lambda}{2p \cdot NA}, \tag{1}$$

where λ is the wavelength of the light in vacuum, p is the pitch of the mask pattern and NA is the numerical aperture of the projection lens (the user-set NA in dry or immersion scanner). This formula can be derived from the grating equation under the symmetry condition where the 0th and $\pm$1st diffraction orders are centered symmetrically about the optical axis. The immersion refractive index n appears in intermediate steps but cancels out in the final expression due to the effective NA relation:

$$NA_{\text{eff}} = \frac{NA}{n}. \tag{2}$$

As such, this expression allows direct calculation of σ based only on optical wavelength, pitch, and NA. On a modern DUV lithography scanner, if the resulting σ exceeds 0.98, it indicates that the required diffraction angle

is beyond the physical pupil boundary, meaning the configuration may not be realizable without a higher NA setting.

3. DIFFUSION MODEL

3.1 Model Overview

Diffusion models have emerged as powerful generative approaches, showing remarkable success in tasks such as conditional image generation. A popular branch is named Latent Diffusion Models (LDM), which operates by progressively transforming random noise into meaningful images through iterative denoising steps in a compact latent space. Specifically, LDMs first encode images into low-dimensional latent representations, significantly reducing computational complexity while preserving essential visual structures. A neural decoder then reconstructs high-quality images from these latent representations.

In this work, we employ the widely adopted Stable Diffusion framework,[24] a mature, open-source LDM known for its reliability and strong generative capability. Leveraging this model, we develop a novel conditional generation pipeline specifically adapted to lithography source optimization, where textual descriptions of illumination parameters (e.g., wavelength, numerical aperture) and layout characteristics (e.g., pitches, pitch angles) guide the generation of optimized illumination source patterns. With the advantages of Stable Diffusion, we aim to efficiently boost the source optimization process through rapid and practical source image generation. In this section, we introduce the technical details of a standard text-to-image task-oriented diffusion model. Its lithographic adaptation and the pipeline will be discussed in the next section.

3.2 Text-to-Image Generation with Diffusion Model

This subsection presents the core technical details underlying the LDM pipeline utilized for conditional text-to-image generation. The process comprises three main components: textual encoding, the latent diffusion process, and image decoding.

Text Encoder The text encoder serves as the interface to convert input textual descriptions into structured numerical embeddings suitable for conditional generation tasks. LDMs typically employ a Transformer-based encoder architecture due to its advantages in capturing semantic relationships and contextual dependencies within textual inputs. The encoding pipeline starts with tokenization, wherein the input text is segmented into subword units. Each subword token is then mapped to an integer index based on a pre-defined vocabulary constructed during model pre-training. The sequence of token indices is then passed through an embedding layer, which projects each token to a high-dimensional continuous vector space. A Transformer-based encoder, consisting of multiple layers of multi-head self-attention and position-wise feedforward modules, processes these embeddings to produce a fixed-length contextualized embedding E. The embedding captures the global semantic content and serves as the conditioning signal for the image generation process.

Diffusion in Latent Space The core generative mechanism in LDMs is the iterative latent diffusion process, which gradually converts a random noise vector into a latent representation of the target image. The diffusion process comprises two stages: a forward diffusion process involving continuous noising, and a reverse diffusion process involving denoising.

In the forward diffusion stage, the model steadily corrupts a given image embedding x_0 into a purely noisy representation x_T over T steps of gradual noising. At each timestep, Gaussian noise is introduced as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t \mathbf{I}), \tag{3}$$

where x_0 is the ground-truth image embedding, $t = 1, \cdots, T$, and β_t is a predetermined noise schedule controlling the incremental amount of noise added at timestep t.

The reverse process reconstructs data from pure noise by learning the reverse transitions. At each step, a neural network parameterized by θ predicts the mean and variance:

$$p_\theta(x_{t-1}|x_t, E) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, E, t), \Sigma_\theta(x_t, E, t)\right), \tag{4}$$

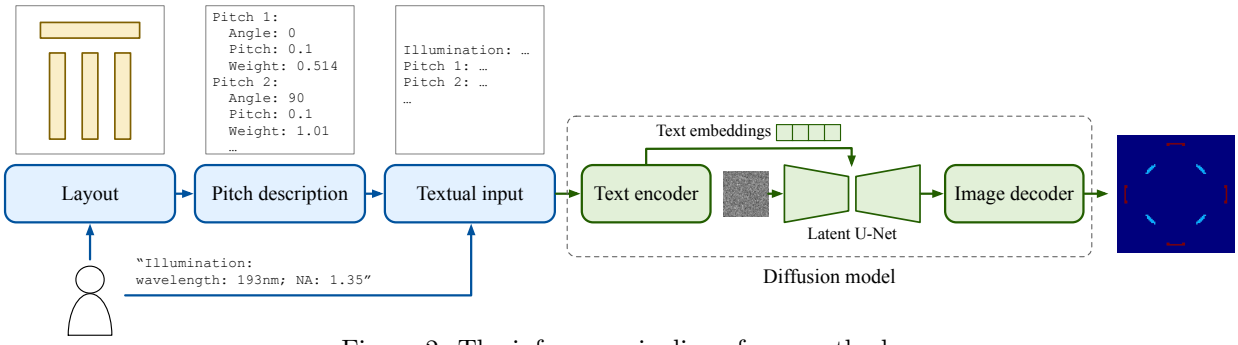


Figure 2: The inference pipeline of our method.

where E denotes the conditional context, and in this case refers to the aforementioned text embedding. The neural network, typically in the form of a U-Net, iteratively denoises x_t by predicting μ_θ and Σ_θ to recover the noised image, with the context E ensuring that the reconstruction aligns with the desired attributes described by the input. After T steps of denoising starting from x_T, the recovered embedding is denoted as $\tilde{x}_0$. It then serves as the input of the decoder to be projected to the pixel space.

Decoding After the reverse diffusion process, the resulting latent representation $\tilde{x}_0$ encodes the semantic and structural content required for generating the final image. A neural network is utilized for mapping the latent representation to pixel space. In Stable Diffusion, the decoder part of a pretrained variational auto-encoder is employed. The decoder is defined as:

$$\hat{x} = \text{Decoder}(\tilde{x}_0), \tag{5}$$

where $\tilde{x}_0$ is the recovered latent from the diffusion process, and $\hat{x}$ is the reconstructed output image.

Training During training, three neural network-based modules are involved: the text encoder, the diffusion network for noise prediction, and the decoder. They jointly minimize a reconstruction objective that measures the discrepancy between the predicted output and the ground truth. Since optimizing in the pixel space brings severe computational complexity, a common training strategy in LDMs is to formulate the loss in the latent space. The decoder is pretrained as part of an autoencoder with its parameters frozen during diffusion model training. In the latent space, the model predicts the noise ϵ added to the ground-truth latent x_0 at each diffusion timestep. The training objective is the mean squared error (MSE) between the predicted and true noise:

$$\mathcal{L}_{\text{latent}}(\theta) = \mathbb{E}_{x_0,\epsilon\sim\mathcal{N}(0,I),t}\left[\|\epsilon - \epsilon_\theta(x_t, t, E)\|^2\right], \tag{6}$$

where ϵ is the actual noise, ϵ_θ is the model's prediction, and E is the conditional embedding. With this formulation, all variables are defined in the latent space. Meanwhile, the text encoder is also pretrained and frozen in most implementations for stability and efficiency. Hence, only the diffusion model parameters are updated in our framework.

4. DIFFUSION MODEL-BASED TRAINING & INFERENCE PIPELINE

This section describes in detail the practical training and inference pipeline for source optimization, including training dataset collection, training framework establishment, and a user-friendly inference pipeline.

4.1 Training Phase

The primary objective of the training phase is to learn a robust mapping from structured textual descriptions of lithographic conditions to corresponding optimized source images. To accomplish this, we prepared a comprehensive training dataset, where each instance encapsulates detailed layout characteristics as well as optical illumination parameters, coupled with the corresponding source pattern image as the ground truth. The data preparation details will be elaborated in Section 5.1. Each instance in the dataset incorporates three components: layout information, illumination conditions, and the corresponding source image as the ground truth. We demonstrate the details of each part as follows:

- Global illumination parameters: wavelength, measured in nanometers; numerical aperture.
- Pitch description: For each identified pitch, we include three critical attributes: (1) orientation angle θ, measured in degrees; (2) pitch value p, in microns; (3) weight w, calculated based on the duty ratio of the corresponding pitch.

Note that the users may provide illumination and process parameters using natural-language descriptions in the inference phase. To ensure consistency and robust generalization, a unified and human-readable serialization template is employed for organizing these parameters in the training stage. Each prompt serves as the conditional input corresponding to a unique source image, which is the direct supervision signal for training the diffusion model.

4.2 Inference Phase

During inference, the trained model generates optimized source images for user-supplied layouts and illumination parameters. As revealed in Fig. 2, the inference pipeline comprises several steps: (1) the user provides the layout for SO, along with the illumination parameters description; (2) pitches, orientations and weights are extracted, formulating a structured and serialized description for the input layout; (3) the illumination parameters provided by the user and the layout description are concatenated and fed into the text encoder for tokenization and image generation.

4.3 Challenge

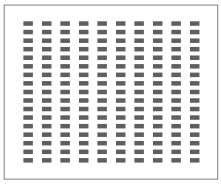
A notable practical challenge encountered is the inherent maximum input length constraint of current text encoders for language models. Such models typically accept a fixed-length token sequence determined by their embedding layers. This limitation becomes critical when dealing with complex layouts, which often result in long serialized input strings due to intensive layout pitches.

To address this issue, we leverage a fundamental lithographic property: the correspondence between layout pitches and their expected illumination conditions reveals an approximate linear superposition characteristic. In other words, the final optimized source image for a complex layout can be effectively approximated as the sum of the individual source data optimized for each pitch. Hence, we design a practical strategy to prevent the input length constraints. We split the full set of layout pitches into multiple mini-batches, each satisfying the input length constraint. Next, each batch is independently serialized into the structured input format, and separately passed through the diffusion model pipeline, producing partially optimized source images. Later on, the sub-images from each batch are linearly superposed and normalized, yielding the final illumination source pattern. This mini-batching and superposition strategy effectively resolves the text length limitation while preserving the integrity of the underlying lithographic rationale.

5. EXPERIMENT

5.1 Dataset

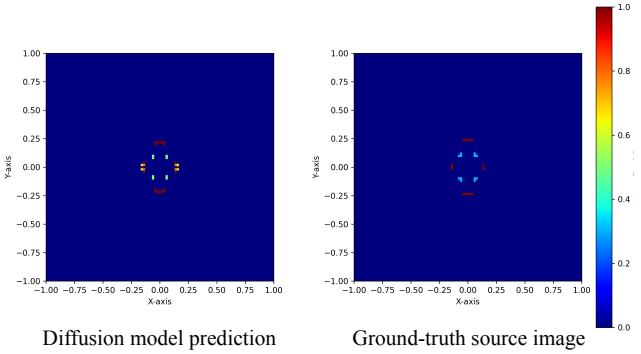
An in-house dataset generation tool has been developed, by implementing sigma optimization in immersion lithography simulations. Starting from stage one, the code generates a set of 1D/2D test patterns based on user-supplied pitch, orientation, and duty ratio-range values. At stage two, the optimized sigma values are calculated using the two-beam imaging formula per pitch per test pattern. The code takes in λ, NA, n, and p, computes σ and checks feasibility, then applies min and max sigma values if needed. Results are output in YAML file containing pitch data including count, orientation and transmittance, and the corresponding sigma result. For stage three, optimized illumination sources are generated for all test patterns. Each optimized source is generated based on the YAML sigma output per test clip from stage two, adding appropriate weight and orientation per pitch for user configured source sectors, then superimpose, normalize and post-process user and scanner constraints for the final illumination source image output.



Testcase with 3 pitches

```
Wavelength 193, NA 1.35;
Pitch 1: Angle: 0, Pitch:
0.500, Weight: 0.542;
Pitch 2: Angle: 90, Pitch:
0.300, Weight: 0.542;
Angle: 59, Pitch: 0.583,
Weight: 0.163;
```

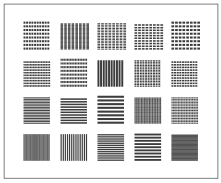
Extracted prompt for diffusion model



Diffusion model prediction

Ground-truth source image

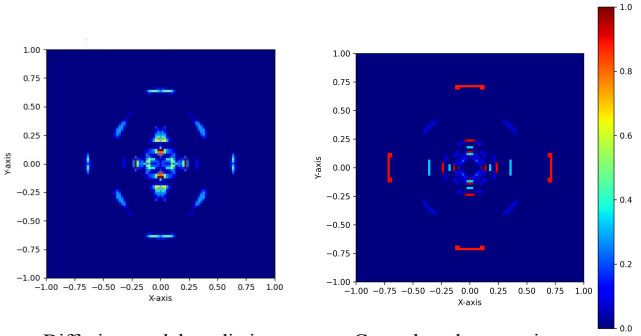
(a)



Testcase with 52 pitches

```
Wavelength 193, NA 1.35;
Pitch 1: Angle: 0,Pitch:
0.2,Weight: 1.02; Pitch 2:
Angle: 0, Pitch: 0.4,
Weight: 1.04;… Pitch 52:
Angle: 45, Pitch: 0.99,
Weight: 0.223;
```

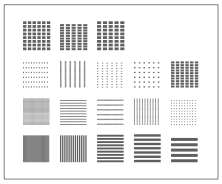
Extracted prompt for diffusion model



Diffusion model prediction

Ground-truth source image

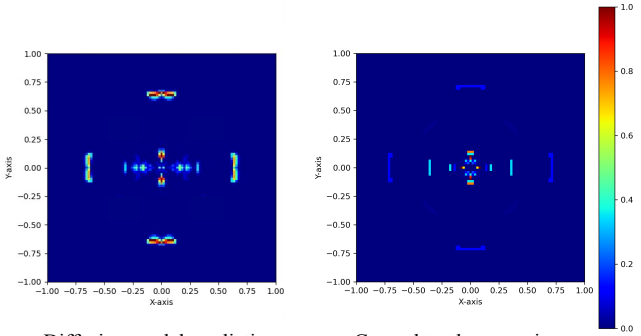
(b)



Testcase with 44 pitches

```
Wavelength 193, NA 1.35;
Pitch 1: Angle: 0,Pitch:
0.2,Weight: 1.02; Pitch
2: Angle: 0, Pitch: 0.4,
Weight: 1.04;… Pitch 44:
Angle: 58, Pitch: 1.304,
Weight: 0.298;
```

Extracted prompt for diffusion model



Diffusion model prediction

Ground-truth source image

(c)

Figure 3: Visualization of three testcases in multiple complexity levels.

5.2 Application Examples

Implementation Details We adopt the open-source implementation of Stable Diffusion v1.4[24] as the backbone for our text-to-image source optimization framework. The pipeline incorporates a pretrained CLIP text encoder,[25] which is utilized as a fixed feature extractor throughout both training and inference. The CLIP text encoder supports input sequences of up to 77 tokens, which corresponds approximately to the text length required for illumination parameters and the description of up to three pitches within a single input prompt. Note that this is only for demonstration purpose, and the text encoder can be extended to larger capacity ones to accommodate longer prompts in practice.

For each training instance, we construct the textual input by concatenating the illumination parameters (wavelength and NA) with a structured list of pitch-specific attributes, as revealed in Fig. 3. When the number of pitches of a layout exceeds three, to address the token limit, we randomly partition the pitch set into multiple batches, each containing at most three pitches. The corresponding illumination sigmas are calculated for each pitch using Equation (1). For each pitch batch, a sub-source image is generated as the supervision signal. We additionally set the sigma angle and depth to fixed values of 20 degrees and 3%. The illumination type is set to quadrupole by default. During model training, the input to the text encoder is formulated on a per-batch basis, and the network aims to predict the corresponding sub-source image.

Visualization We present three test cases with varying layout complexity, as revealed in Fig. 3. For each test case, we display the layout pattern, the corresponding extracted textual prompt used as model input, and a visual comparison between the predicted source image generated by the diffusion model and the ground-truth source image. For the two complicated cases, since the prompt length exceeds the limit, we split them every three pitches, and sum the sub-images as in the training phase. In all cases, the source images are plotted with their spatial domain normalized to $[-1, 1]$ along both axes, and the numerical range of each image is normalized to $[0, 1]$ for consistent visualization. The close correspondence between the diffusion model predictions and the physical ground-truth across different complexities demonstrates the generalization capability of our framework, as well as its practical potential for boosting the SO process.

6. DISCUSSION

Our work tackles a key production challenge—resolving lithographic hot spots as quickly as possible. With our framework, engineers can simply ask the SO tool in natural language, for example: "Given this NA and exposure conditions, how can we improve DOF for critical pitches around 60 nm and 115 nm on this reticle?" The tool then instantly produces an optimized illumination profile that can be overlaid directly onto the existing source map.

Several enhancements could further extend the capabilities of our proposed framework, mainly involving: (1) semantic data augmentation for better generalization, and (2) overlong input solutions beyond pitch decomposition.

One potential improvement in our framework is enhancing the model's robustness to linguistic variability in user-provided inputs. In practical applications, users may describe identical illumination and source type information using a wide variety of phrasings and grammar structures. This diversity challenges the model's generalization. To address this, a promising strategy is semantic data augmentation. This approach involves expanding the training dataset by generating paraphrased versions of input descriptions that express the same underlying process parameters and source conditions. For example, the description "Wavelength is 193nm, NA is 1.35" can be augmented with variants such as "The numerical aperture is 1.35 and the exposure wavelength is 193 nm," or "Using a 193-nanometer wavelength with NA=1.35." Paraphrasing can be achieved by setting manual templates, using paraphrase generation models, etc. Applying semantic data augmentation enhances the ability of the text encoder—and thus the overall system—to generalize to diverse user input styles.

Another potential challenge lies in addressing overlong textual input. While our current approach to this problem leverages the near-linear superposability of pitch-specific source patterns in SO, such task-specific properties may not hold in more complex or nonlinear applications. For tasks in which the model's output depends on broader contextual information or nonlinear I/O relationships, a potential solution is to employ strategies inspired by statistical distribution estimation. Specifically, one can divide the overlong input into multiple clips, and apply random sampling with replacement to construct a diverse training dataset. By repeated sampling, it is possible to approximate the distributional characteristics of the full input. This Monte Carlo-style approach allows the model to aggregate information from various local contexts throughout multiple training or inference passes, thereby forming an empirical estimate of the global data distribution.

Acknowledgement

The authors would like to acknowledge our intern student, Alexander H. Chen (Lake Oswego High School), for his valuable contributions in setting up test cases, debugging the code, running simulations and generating

data. We also thank Hong Chen, Vice President of the Research Institute at GWX Technology, for her insightful consultation and constructive feedback leading to the creation of this work.

REFERENCES

[1] Erdmann, A., Fühner, T., Shao, F., and Evanschitzky, P., "Lithography simulation: modeling techniques and selected applications," in [*Modeling Aspects in Optical Metrology II*], **7390**, SPIE (2009).

[2] Melville, D. et al., "Demonstrating the benefits of source-mask optimization and enabling technologies through experiment and simulations," in [*Proc. SPIE*], **7274**, 72740T (2009).

[3] EE Times, "What is Source-Mask Optimization?," (2014). Retrieved from https://www.eetimes.com/what-is-source-mask-optimization/.

[4] Semiconductor Engineering, "Design Compliant Source Mask Optimization (SMO)," (2014). Retrieved from https://semiengineering.com/design-compliant-source-mask-optimization-smo/.

[5] Jia, N. and Lam, E. Y., "Pixelated source mask optimization for process robustness in optical lithography," *Opt. Express* **19**(20), 19376–19385 (2011).

[6] Ma, X., Li, Y., Guo, X., Dong, L., and Arce, G. R., "Vectorial mask optimization methods for robust optical lithography," *J. Micro/Nanolith. MEMS MOEMS* **11**(4), 043008 (2012).

[7] Li, J., Liu, S., and Lam, E. Y., "Efficient source and mask optimization with augmented Lagrangian methods in optical lithography," *Opt. Express* **21**(7), 8076–8088 (2013).

[8] Anonymous, "Fast freeform source and mask co-optimization method." US Patent US10592633B2.

[9] Zou, Y., "Evaluation of Lithographic Benefits of using ILT Techniques for 22nm-node," in [*Proc. SPIE*], **7640**, 764006 (2010).

[10] Shen, Y., Jia, N., Wong, N., and Lam, E. Y., "Robust level-set-based inverse lithography," *J. Opt. Soc. Am. A* **29**(11), 2319–2328 (2012).

[11] Shen, Y., Wong, N., and Lam, E. Y., "Level-set-based inverse lithography for photomask synthesis," *Opt. Express* **20**(4), 3670–3682 (2012).

[12] Rosenbluth, A. et al., "Intensive optimization of masks and sources for 22-nm lithography," in [*Proc. SPIE*], **7274**, 724209 (2009).

[13] Xiao, G., "Source Optimization and Mask Design to Minimize MEEF in Low k1 Lithography," in [*Proc. SPIE*], **7274**, 72741C (2009).

[14] Lafferty, N. et al., "Full-flow RET creation, comparison, and selection," in [*Proc. SPIE*], **9235**, 92351Z (2014).

[15] Lafferty, N. V. et al., "RET selection on state-of-the-art NAND flash," in [*Proc. SPIE*], **9426**, 94260L (2015).

[16] Zhang, D. et al., "Source mask optimization methodology (SMO) and application to real full chip optical proximity correction," in [*Proc. SPIE*], **7274**, 72742B (2009).

[17] Chen, G. et al., "Physics-Informed Optical Kernel Regression Using Complex-Valued Neural Fields," (2023). arXiv:2303.08435.

[18] Yang, H. et al., "Large Scale Mask Optimization Via Convolutional Fourier Neural Operator and Litho-Guided Self Training," (2022). arXiv:2207.04056.

[19] Chen, G. et al., "DevelSet: Deep Neural Level Set for Instant Mask Optimization," (2023). arXiv:2303.12529.

[20] Chen, G. et al., "DAMO: Deep Agile Mask Optimization for Full Chip Scale," (2020). arXiv:2008.00806.

[21] Wu, C.-W. et al., "Freeform source optimization for improving litho-performance of warm spots," in [*Proc. SPIE*], **8166**, 81663C (2011).

[22] Yu, J.-C., Yu, P., and Chao, H.-Y., "Source optimization incorporating margin image average with conjugate gradient method," in [*Proc. SPIE*], **8166**, 81662B (2011).

[23] Levinson, H. J., [*Principles of Lithography, 3rd Ed.*], SPIE Press (2010).

[24] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B., "High-resolution image synthesis with latent diffusion models," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 10684–10695 (June 2022).

[25] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I., "Learning Transferable Visual Models From Natural Language Supervision," in [*Proceedings of the 38th International Conference on Machine Learning (ICML)*], *PMLR* **139**, 8748–8763 (2021). arXiv:2103.00020.