Supplementary Information for:
Full-chip EUV curvilinear mask optimization

Pinxuan He[1], Jiamin Liu[1*], Honggang Gu[1], Song Zhang[1,3], Qi Xia[1], Hao Jiang[1], Shiyuan Liu[1,2*]

[1]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.
[2]Optics Valley Laboratory, Wuhan 430074, China.
[3]Yuwei Optica Co., Ltd, Wuhan 430074, China.

*Corresponding author(s). E-mail(s): jiaminliu@hust.edu.cn; shyliu@hust.edu.cn;

Surrogate models

Two AI models, with the same architectures, are used to perform inference for x-polarized and y-polarized incident field, respectively. The AI models take plane wave parameters α, β, and pattern to predict the amplitude perturbation ΔAmp and phase perturbation ΔPh. As sketched in Fig. S1, the network consists of one encoder for pattern, one fully connected network (FCN) for incident information, and one decoder for outputs. The numbers next to the vectors indicate their element counts, while the numbers above the three dimensional data represent the number of channels. The number of channels in each convolution block is the same and is shown in the upper-right corner. Arrows in different colors indicate different operations.

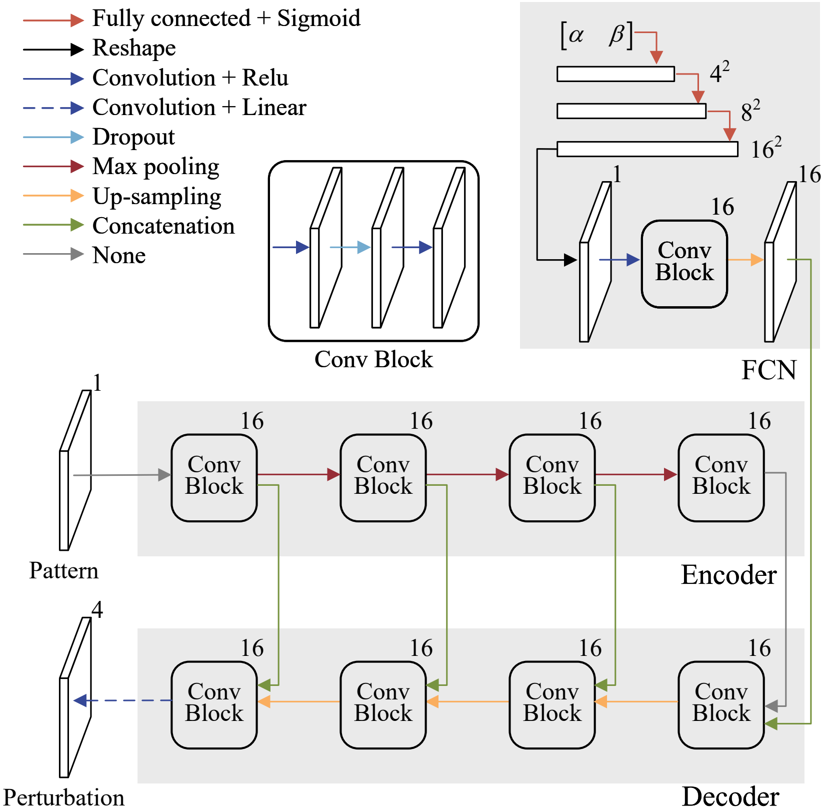

Fig. S1: Schematic overview of the encoder and decoder architectures. The numbers next to the vectors indicate their element counts, while the numbers above the three dimensional data represent the number of channels. The number of channels in each convolution block is the same and is shown in the upper-right corner. Arrows in different colors indicate different operations.

The encoder-decoder design, along with the long-skip connections via concatenation, is a classical feature of the U-Net architecture. The filters size in all convolutions is 3×3. The dropout layers with a factor of 0.01 is incorporated into the network to mitigate overfitting. The pooling/upsampling layers use 2×2 filters. The number of channels used by the convolution encoder block is 16.

The integration of the FCN is inspired by a tunable U-net design. The input vector, which has two elements, is sequentially connected to vectors of size 4^2, 8^2, and 16^2. The 16^2 vector is subsequently reshaped into a 16×16 array, which is followed by several operations to generate multiple channels and adjust them to the correct size for concatenation with the encoder's output. The inference of the models is accelerated by TensorFlow RT 8.6.1.

Model preparation

The AI models are designed to predict the near field for any given pattern, critical dimension (CD), and source points. Although there are lots of differences between the DUV and EUV lithography, it is acceptable to use a pattern library from DUV mask optimization [1] to prepare the dataset for EUV near field predictions. The pattern library contains 16,472 patterns, and these patterns were clipped to retain only the central 63.8% of its length, minimizing computations over empty areas. The typical patterns after clipping are shown in Fig. S2. The models' accuracy will not be affected because they are surrogate models rather than end-to-end models. The source point will be chosen from the illumination pupil. The corresponding CD on the wafer, without consideration of thick mask effect and diffraction-limited feature of projection system, ranges from 10 to 40 nm. The training data was generated for each polarization through 3,000 reference simulations using random patterns, CD, and source points. An additional 10% of the data was reserved as a validation set.

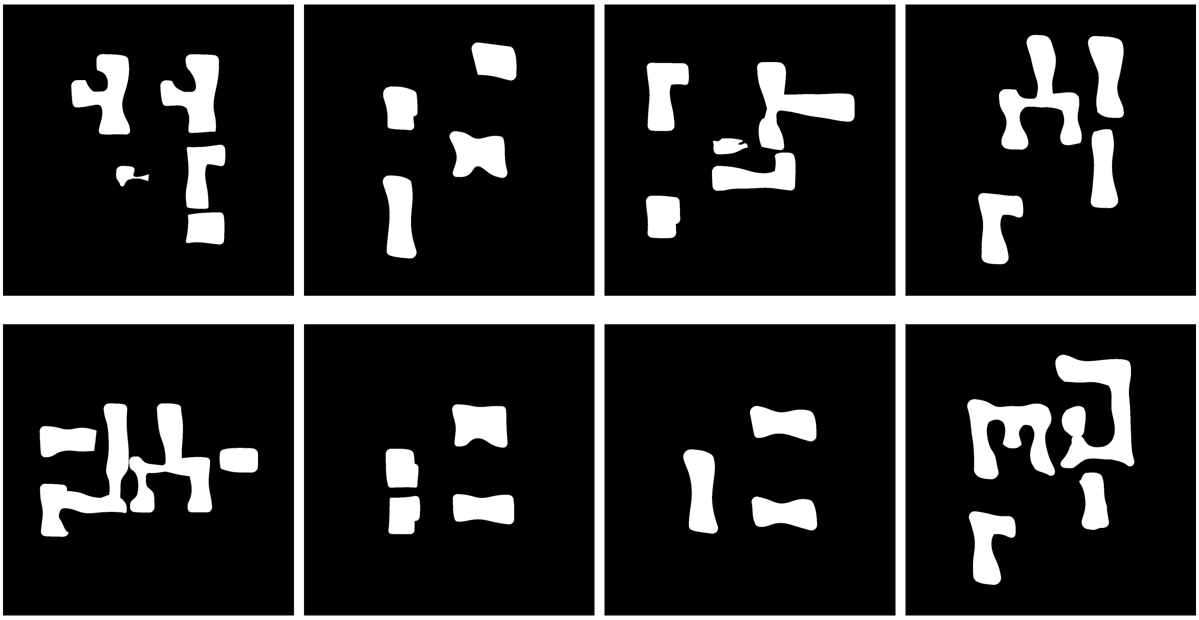

Fig. S2: Typical pattern from DUV mask optimization [1] used for data generation.

Before further processing the reference field, reference simulations without pattern are conducted for each source points to get the reference amplitude and phase. The simulations' area can be small because of the absence of the pattern. Thus, preparing the reference amplitude and phase does not require much time.

The amplitude and phase are extracted from the complex-valued field. For the amplitude, the amplitude perturbation is obtained by subtracting the reference amplitude. The phase is unwrapped [2] before the subtraction. Finally, the amplitude and phase perturbation are tiled into patches.

Predictions at the edges of the patches are expected to be inaccurate due to the lack of neighboring information. Therefore, a window function, as shown in Eq. 1 is applied to both the prediction and reference before calculating the loss.

$$w(x) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi x}{\frac{L}{2}-1}\right) & 0 \leq x < \frac{L}{4} \\ 1 & \frac{L}{4} \leq x < \frac{3L}{4} \\ 0.5 - 0.5\cos\left(\frac{2\pi(L-1-x)}{\frac{L}{2}-1}\right) & \frac{3L}{4} \leq x < L \end{cases} \tag{1}$$

where L is the length of the data, set to 256 in this work. The loss function consists of amplitude and phase parts. The amplitude part is the mean square error (MSE) of the processed prediction and reference.

$$loss_{\mathrm{Am}} = \frac{1}{N}\sum\left(\Delta Am_{\mathrm{true}} - \Delta Am_{\mathrm{pred}}\right)^2 \tag{2}$$

The phase part contains the MSE for the x or y polarization and the weighted MSE for z polarization.

$$loss_{\mathrm{Ph}} = \frac{1}{N}\sum\left(\Delta Ph^{\mathrm{xy}}_{\mathrm{true}} - \Delta Ph^{\mathrm{xy}}_{\mathrm{pred}}\right)^2 + \frac{1}{N}\sum \Delta Am^{\mathrm{z}}_{\mathrm{true}} \cdot \left(\Delta Ph^{\mathrm{z}}_{\mathrm{true}} - \Delta Ph^{\mathrm{z}}_{\mathrm{pred}}\right)^2 \tag{3}$$

The final loss is the weighted combination of the amplitude and phase parts.

$$loss = loss_{\mathrm{Am}} + loss_{\mathrm{Ph}}/(5.5)^2 \tag{4}$$

The models are trained using the RMSprop optimizer with a learning rate of 0.001 and a batch size of 32. The model are trained for 20 epochs.

After the training, the model is further calibrated using the reference model. First, the aerial images of a calibration pattern under all source points are calculated. The calibration pattern is shown in Fig. S3. Then, the same result are obtained by the AI models. Calibration coefficient for the j^{th} source points is defined as

$$\mathrm{Coe}_j = \arg\min_s \left(\sum_j w(I_j) \cdot \left|I_{\mathrm{ref},j} - s \cdot I_j\right|\right) \tag{5}$$

where $w(x) = \exp\left(-\left(\frac{x-\mathrm{Th}}{0.2}\right)^2\right)$. Th is the threshold of the resist model. Essentially, the calibration coefficients are global scaling factors for the intensity. These coefficients will be applied to the aerial image for each source point.

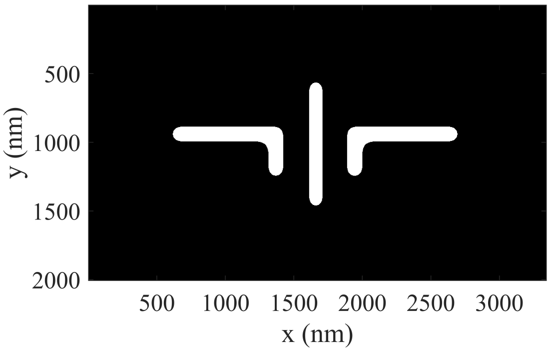

Fig. S3: The mask pattern used to obtain the calibration coefficients. The coordinates are in mask scale.

Beam propagation method

Beam propagation method (BPM) is a powerful tool for simulation of propagation of light [3–5]. It can only be applied to situation where the reflected field can be neglected and the variation of refractive index in medium is small. Besides, the BPM assumes paraxiality and small propagation step. Given these restrictions, the BPM still has wide application in many areas. For EUV mask simulation, these restrictions are relaxed because the refractive index of the typical absorber in an EUV mask is close to that of the surrounding medium [6].

One implementation of the BPM is demonstrated in this section, shown in Algorithm 1. F_{2D} and F^{-1}_{2D} denote the 2D forward and inverse Fourier transform operators, respectively. Note that there are different implementations of the BPM [3, 7], and each of them can be applied because the BPM is inherently an approximation technique.

Algorithm 1 Beam Propagation Method

```
Require: $E_{\mathrm{sour}}(x, y)$, $n(x, y; N)$
Ensure: $E_{\mathrm{out}}(x, y)$
Transform the source field into frequency domain:
$E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow F_{\mathrm{2D}}\left\{E_{\mathrm{sour}}(x, y)\right\}$
Propagate forward half of $\Delta z$ in slice 1:
$E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \cdot p\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right)$
for $j = 2$ to $N_z - 1$ do
    Propagate forward half of $\Delta z$ in slice $j$:
    $E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \cdot p\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right)$
    Transform $E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}})$ into time domain:
    $E_{\mathrm{temp}}(x, y) \leftarrow F_{\mathrm{2D}}^{-1}\left\{E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}})\right\}$
    Apply the scattering function of slice $j$:
    $E_{\mathrm{temp}}(x, y) \leftarrow E_{\mathrm{temp}}(x, y) \cdot s\left(x, y; j\right)$
    Transform $E_{\mathrm{temp}}(x, y)$ into frequency domain:
    $E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow F_{\mathrm{2D}}\left\{E_{\mathrm{temp}}(x, y)\right\}$
    Propagate forward half of $\Delta z$ in slice $j$ again:
    $E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \cdot p\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right)$
end for
Propagate forward half of $\Delta z$ in slice $N_z$:
$E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \leftarrow E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}}) \cdot p\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right)$
Return the results:
$E_{\mathrm{out}}(x, y) \leftarrow F_{\mathrm{2D}}^{-1}\left\{E_{\mathrm{temp}}(k_{\mathrm{x}}, k_{\mathrm{y}})\right\}$
```

First, the simulation domain is divided into slices with fixed intervals along the propagation direction. Then, the field is converted in to frequency domain and apply a propagation function

$$p\left(k_{\mathrm{x}}, k_{\mathrm{y}}\right) = \exp\left(1\mathrm{i} \times 0.5 \times \Delta z \times \sqrt{\left(k_0 n_{\mathrm{bg}}\right)^2 - k_{\mathrm{x}}^2 - k_{\mathrm{y}}^2}\right) \tag{6}$$

where k_0 denotes the free-space wavenumber; Δz is the propagation step; n_{bg} is the background refractive index. Then, the propagation in frequency domain and scattering in space domain are conducted alternately for each slice. The scattering function is defined as

$$s\left(x, y; N_z\right) = \exp(ik_0(n(x, y; N_z) - n_{\mathrm{bg}})\Delta z) \tag{7}$$

where N_z is an index representing the z position of the calculated slice, and $n(x, y; N_z)$ is the refractive index of the slice indexed by N_z. After the loop, the field is further propagated by half of Δz. Finally, the result can be obtained by inverse Fourier transform.

Accuracy validation

Although the accuracy of the proposed framework is independent of the reference model, which can be replaced with any other qualified mask model, it is still necessary to report the accuracy of both the reference mask model and the projection model for justifying the following accuracy validation. The used resist model is simple but widely applied[8–10], which is sufficient for this proof-of-concept study focusing primarily on the mask model.

The near-field and aerial image of a line/space pattern were obtained and compared against results from a commercial tool, as shown in Fig. S4. The pattern had a pitch of 79.875 nm and a CD of 39.937 nm on the wafer. The illumination is a single y-polarized point source at an incidence angle of 5.98°. The Ta_6N_4 absorber had a thickness of 49.5 nm, and the multilayer reflector consisted of ideal Mo/Si bilayers[11]. All other unspecified parameters follow the specifications in the main text Methods section. The near-field results from both methods show good agreement in amplitude and phase. These minor differences in the near fields translate to a maximum deviation of only 0.034 in the aerial image.

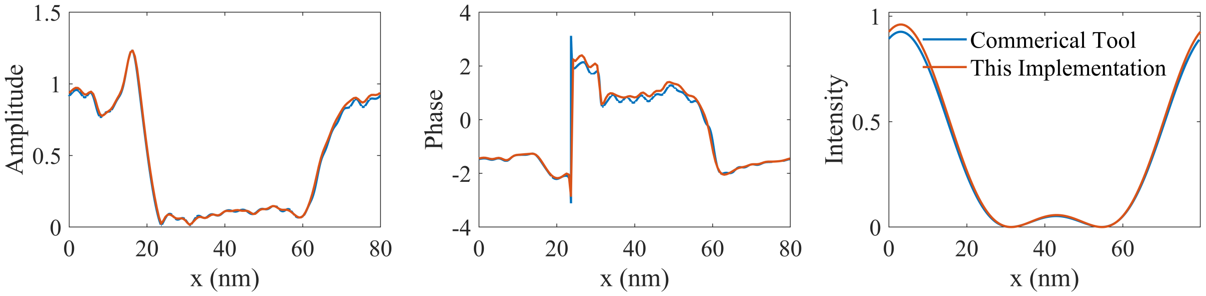

Fig. S4: Comparison between our method and a commercial tool. A line/space pattern was simulated. From left to right are the mask near-field amplitude, phase, and aerial image. The color legend is shared across all plots.

Next, we focus on comparing the reference model with the surrogate model. A comparison between the proposed framework and the reference one in terms of the near field, aerial image, and gradient is demonstrated in Fig. S5. The test case is a typical mask pattern with local curvilinear features, as shown in Fig. S5e. The minimum CD in the mask pattern is 118.125 nm, measured at y = 2100 nm, x = [2120, 2250] nm. Given the demagnification factor of 4 in the projection system, the corresponding CD on the wafer is 29.531 nm without accounting for the diffraction-limited feature and thick mask effect. Note that gradient calculation with the adjoint method under PCI is computationally expensive. The typical near field and gradient are presented for only one TE-polarized source point, with a wave vector of $[0.1055, -0.0073, 0.9944]$.

Although the direct output of the surrogate models are perturbations, the near field is presented in terms of the amplitude and phase perturbation for clarity. The amplitudes from the reference and surrogate models are presented in Fig. S5a and S5b, respectively. The predicted amplitude demonstrates excellent agreement with reference data. Comparable consistency can be observed in the phase perturbation comparison shown in Fig. S5c and Fig. S5d.

The aerial image under the PCI, specifically under 332 source points, is calculated using both the reference and surrogate models. The reference aerial image and the differences between the two results are shown in Fig. S5f and S5g, respectively. The differences are primarily concentrated in the center of the pattern, while the surrogate model shows high accuracy at the edges. It could be attributed to the fact that the U-Net was originally designed for medical image segmentation and is highly sensitive to edges [12]. In MO, the prediction accuracy at the edges of the pattern is more important than that at the center of the pattern. The corresponding evaluation criterion in mask optimization is the average edge placement error, as shown below

$$F(I) = \frac{\|resist(I) - C_{\text{target}}\|_2^2}{P} \tag{8}$$

where I represents the aerial image. The resist model has a threshold effect where the material activates above a limit and stays inert below, making the predictions of the field edge dominant [13]. The resultant contour is compared with the target contour, and the deviation is averaged over the perimeter P of the target pattern. The notation $\|\cdot\|_2^2$ represents the square of l_2-norm. As shown in Fig. S5h, the vertical and horizontal slices at the locations of the maximum differences also validate the accuracy of the surrogate model.

The approximated gradient is compared with the gradient from the adjoint method in Figs. S5i and S5j. Although their overall distributions are similar, local differences are expected

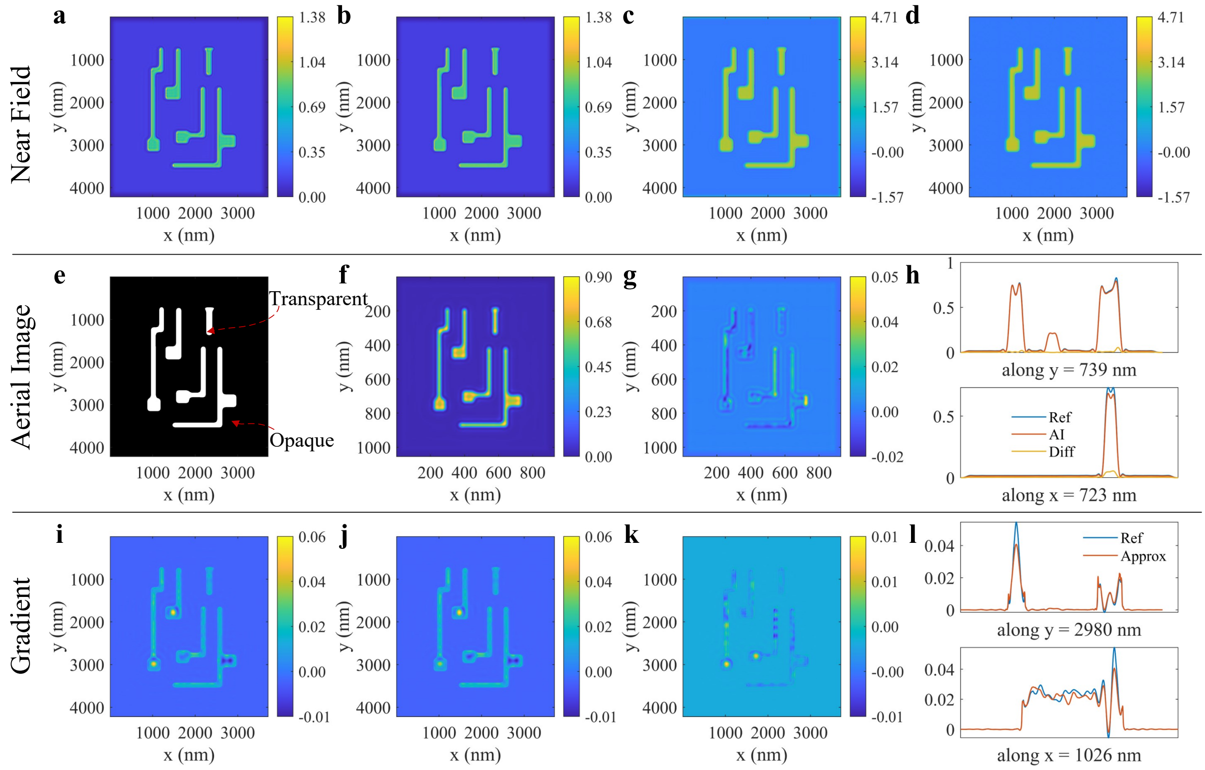

Fig. S5: Accuracy validation of the surrogate models and the approximated gradient. (e) The mask pattern for the accuracy evaluation. The black region is covered by the absorber, and vice versa. The amplitude is calculated from (a) the reference model and (b) the surrogate models. The phase perturbation is calculated from (c) the reference model and (d) the surrogate models. (f) The aerial image from the reference model. (g) The differences between the aerial images from the reference and surrogate models. (h) The horizontal and vertical cross-sections of the aerial images at the locations of maximum differences. The differences are also plotted. (i) The reference gradient from the adjoint method. (j) The approximated gradient. (k) The differences between the gradients from the adjoint method and approximated method. (l) The horizontal and vertical slices of the gradients at the locations of maximum differences. The cross-section positions are annotated in the subfigures.

due to the adoption of the scalar simulation method. The difference between the gradients is presented in Fig. S5k. The vertical and horizontal slices at the locations of the maximum differences are also provided in Fig. S5l. Although the relative error is obvious, reaching up to 33%, the overall trend remains consistent. Although moderate gradient errors may alter the optimization trajectory and lead to variations in the final cost value, they are acceptable as long as the optimized target, namely the wafer pattern, remains satisfactory. Moreover, these errors have no impact on the final accuracy, which is determined by the forward model.

Cost convergence

Here is the cost convergence curve of the large-scale mask optimization example, where the last stage is skipped.

MO example under Y-Dipole illumination

This supplementary example demonstrates mask optimization under Y-dipole illumination. The source, comprising 336 points, is shown in Fig. S7a. On the wafer side, the CD is still 19.41 nm and the area is 0.77 μm^2 (0.89 μm$\times$ 0.86 μm). The total runtime is approximately 764.3 seconds. The dipole illumination comprises two clusters of source points. On the first stage, a single source

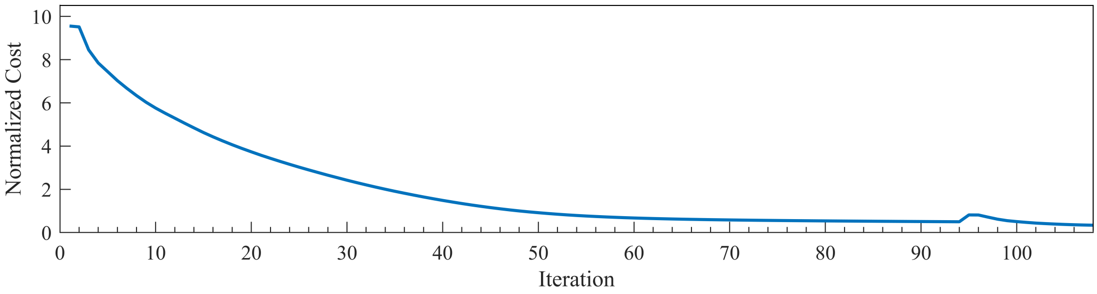

Fig. S6: The cost convergence curve of the large-scale mask optimization example.

point was placed at the centroid of each cluster. Sub-resolution assist features (SRAFs) were also introduced at this stage. With SRAFs, the optimization converged to a final cost of 1.01. Without them, it converged at 3.43, leaving the two outermost bars on the wafer distorted.

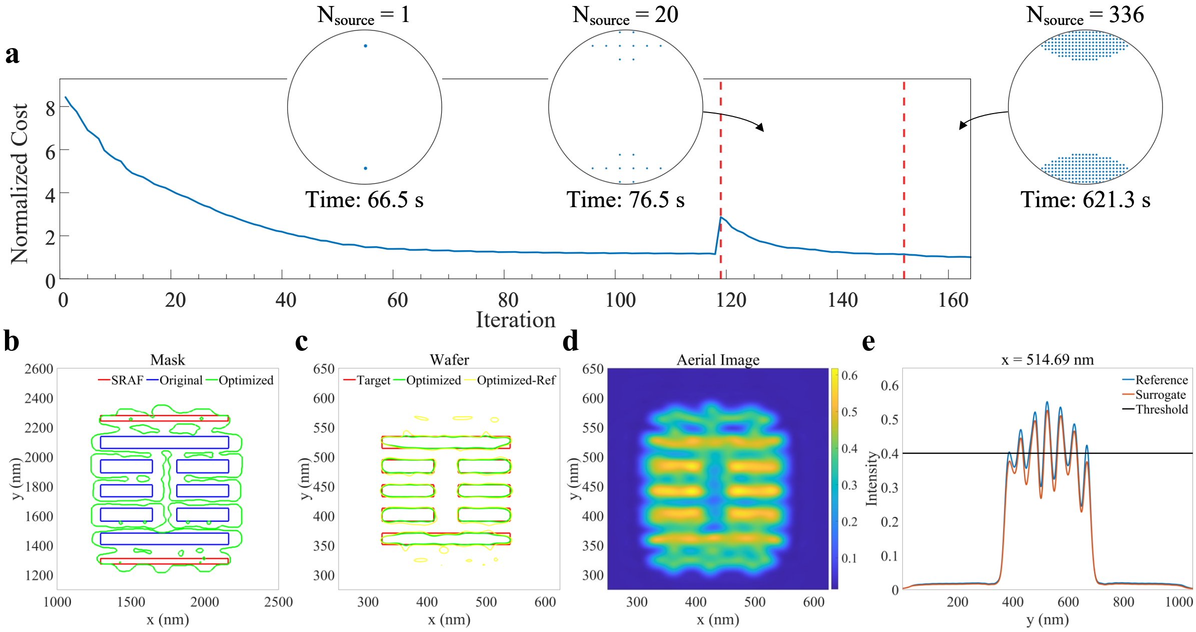

Fig. S7: Mask optimization example on a lines pattern with the critical dimension of 19.41 nm. (a) The convergence of the normalized cost during the optimization. The optimization consists of three stages, each characterized by a different number of source points N_{source}. (b) Comparison between the original and optimized mask pattern contours. (c) Comparison between the target, optimized, and verified wafer pattern contours. The verified wafer contour is obtained using the reference model. (d) The aerial image after optimization from the reference model. (e) The cross-section of (d) along $x = 514.69$ nm.

As seen in Fig. S7c, the predicted wafer pattern deviates from the reference. This discrepancy can be attributed to a failed optimization, rather than to a flaw in the models. As shown in Fig. S7d, the intensity in the SRAFs regions is near to the resist threshold of 0.4, resulting in poor exposure latitude. For clarity, a cross-section comparison of the aerial image is shown in Fig. S7e. Although there are differences at the top and bottom, they generally match well near the threshold, which is acceptable because the edge placement is of primary importance in lithography as discussed above. Future improvements could incorporate image contrast into the cost function, though this is beyond the current scope. In conclusion, the observed wafer discrepancy is directly attributable to near-threshold intensity, rather than a flaw in the proposed method, which remains valid and robust.

Like other advanced illuminations, dipole sources are typically used for specific, oriented patterns. This demonstration uses a general library of random patterns due to the lack of a dedicated, matched pattern set. The method demonstrates its robustness by functioning effectively without application-specific tuning, though a matched library would yield better results.

References

[1] Zheng, S., Yang, H., Zhu, B., Yu, B. & Wong, M. *Lithobench: Benchmarking ai computational lithography for semiconductor manufacturing. Advances in Neural Information Processing Systems*, Vol. 36, 30243–30254 (Curran Associates, Inc., 2023).

[2] Herráez, M. A., Burton, D. R., Lalor, M. J. & Gdeisat, M. A. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Appl. Opt.* **41**, 7437–7444 (2002).

[3] Zhang, Y., Cui, Z., Ji, X., Wang, H. & Dai, Q. 3D fourier ptychographic microscopy based on the beam propagation method and time-reversal scheme. *IEEE Access* **7**, 129402–129410 (2019).

[4] Glaser, A. K., Chen, Y. & Liu, J. T. C. Fractal propagation method enables realistic optical microscopy simulations in biological tissues. *Optica* **3**, 861–869 (2016).

[5] Pedrola, G. L. *Beam propagation method for design of optical waveguide devices* (John Wiley & Sons, 2015).

[6] Takahata, Y., Kovalevich, T., Simone, D. D., Tanaka, Y. & Philipsen, V. Study of mask error enhancement factor improvement with low-n absorber extreme ultraviolet lithography mask. *Journal of Micro/Nanopatterning, Materials, and Metrology* **23**, 044401 (2024).

[7] Guo, J. Z. Y. & Cerrina, F. Modeling x-ray proximity lithography. *IBM Journal of Research and Development* **37**, 331–350 (1993).

[8] Jiang, B. *et al. Neural-ilt: migrating ilt to neural networks for mask printability and complexity co-optimization. Proceedings of the 39th International Conference on Computer-Aided Design*, ICCAD '20 (Association for Computing Machinery, New York, NY, USA, 2020).

[9] Ma, X., Wang, Z., Chen, X., Li, Y. & Arce, G. R. Gradient-based source mask optimization for extreme ultraviolet lithography. *IEEE Transactions on Computational Imaging* **5**, 120–135 (2019).

[10] Poonawala, A. & Milanfar, P. Mask design for optical microlithography—an inverse imaging problem. *IEEE Transactions on Image Processing* **16**, 774–788 (2007).

[11] Philipsen, V. *et al. Actinic characterization and modeling of the EUV mask stack. 29th European Mask and Lithography Conference*, Vol. 8886, 88860B. International Society for Optics and Photonics (SPIE, 2013).

[12] Ronneberger, O., Fischer, P. & Brox, T. *U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).

[13] Lawson, R. A. & Robinson, A. P. in *Chapter 1 - overview of materials and processes for lithography Materials and Processes for Next Generation Lithography*, Vol. 11 of *Frontiers of Nanoscience* 1–90 (Elsevier, 2016).