

Physics-Informed Neural Operator for Warm-Starting Background-Decomposed and Preconditioned PSFD: Enabling Scalable 3-D EUV Mask Simulation

Doyun Kim*and Werner Gillijns[†]

IMEC, Kapeldreef 75, 3001 Leuven, Belgium

Abstract

We present a physics-informed neural operator (PINO) trained with pseudo-spectral frequency-domain (PSFD) equations for electromagnetic (EM) scattering problems in EUV lithography. The Fourier neural operator is factorized into a two-dimensional lateral (xy) branch and a one-dimensional axial (z) branch and is trained self-consistently with background decomposition. Thus, the full-vector coupling between the mask and the multilayer response is retained without invoking a finite-order Born approximation. In this way, the computational domain size is significantly reduced, thereby lowering the computational cost. The PINO is trained on approximately 16,000 mask designs from the LithoBench library sampled randomly at each training iteration without using precomputed EM field solutions. The PINO surrogate model yields predictions with a mean absolute error of about 7×10^{-3} for the scattered intensity of held-out mask patterns relative to the reference PSFD solution. Combined with spectral damping, the PINO warm-start initialization accelerates the background-decomposed PSFD solver on finer discretizations.

1 Introduction

Simulation of electromagnetic fields in three-dimensional structures is a fundamental challenge in computational electromagnetics (EM), with applications spanning photonics, optical metrology, and semiconductor manufacturing[1, 2]. In particular, extreme ultraviolet (EUV) lithography has profoundly transformed semiconductor microlithography. Achieving sub-nanometer control of printed dimensions over large mask layouts requires electromagnetic simulations that are both highly accurate and computationally efficient, although these objectives are difficult to satisfy simultaneously. Traditional rigorous electromagnetic solvers, such as finite-difference time-domain (FDTD) and rigorous coupled-wave analysis (RCWA), while accurate, often struggle to meet the computational demands of simulating large-scale EUV mask structures due to their high computational cost and memory requirements [3, 4]. For this reason, considerable effort has been devoted to developing more efficient computational methods that retain the essential physics while significantly reducing runtime.

*doyun.kim@imec.be

[†]werner.gillijns@imec.be

A common high-throughput OPC workflow combines a Hopkins-type partially coherent imaging formulation with a thin-mask complex mask function, thereby neglecting the volumetric electromagnetic interactions within the physical mask stack [5, 6]. While this combined approximation is computationally efficient, the thin-mask representation cannot fully capture the vectorial near-field interactions and internal coupling that arise in realistic EUV mask structures, leading to inaccuracies in the predicted aerial images and subsequent lithographic performance [7, 8]. In EUV lithography, oblique illumination of the reflective mask and the finite thickness of the absorber stack give rise to pronounced mask 3D (M3D) effects, including orientation-dependent shadowing and phase distortions [9, 10].

To overcome these limitations, several computationally efficient EM modeling approaches have been proposed to approximate mask near fields at a lower cost than full rigorous solvers. Kernel-based compact 3-D mask models approximate rigorous mask near fields by convolving calibrated spatial-domain filters with thin-mask transmission functions, where the filters are calibrated against rigorous electromagnetic simulations of calibration structures [11]. The domain decomposition method (DDM) decomposes complex mask layouts into primitive sub-patterns, computes their EM near-fields rigorously, and assembles the full mask near-field via linear superposition [12]. Data-driven approaches have also been explored, in which machine learning surrogate models are trained on rigorously computed EM field datasets and applied for fast near-field inference [13, 14].

In particular, physics-informed neural networks and neural operators (PINNs and PINOs) have emerged as powerful tools for solving PDEs and integral equations by embedding the underlying physical laws into the learning process [15, 16]. Notable advances in PINNs and PINOs have demonstrated their ability to approximate solutions and solution operators constrained by governing equations, including Helmholtz-type formulations relevant to EUV mask diffraction [17, 18, 19], and to provide efficient surrogates for complex scattering problems [20, 21]. Nevertheless, scalable three-dimensional EUV mask simulation remains computationally demanding despite advances in both rigorous solvers and approximate mask-modeling approaches [22, 23].

Recently, we presented a pseudo-spectral frequency-domain (PSFD) method featuring a background decomposition approach, where the simulation domain is significantly reduced to only the mask region, and introduced preconditioning techniques that further accelerate convergence [24]. In this work, that framework is extended to a PINO based on an anisotropically factorized Fourier neural operator inspired by F-FNO [25] and trained with background decomposition. The resulting surrogate infers 3-D vectorial electric fields for held-out EUV mask geometries without relying on precomputed EM field solutions. We then examine its transfer to a finer lateral discretization of the training domain, without retraining, and assess its use as a warm-start initializer for the background-decomposed PSFD solver with spectral preconditioning.

2 Methodology

2.1 Factorized Fourier Neural Operator for EM scattering problems

A neural operator is a deep learning architecture designed to learn mappings between infinite-dimensional function spaces, making it uniquely capable of approximating the solution operators of partial differential equations (PDEs) [26, 16, 27]. Unlike conventional neural networks, a neural operator processes continuous functions as inputs and outputs. In its foundational formu-

lation, each operator layer typically maps a hidden state via [27]

$$\mathcal{N}(u)(x) = \sigma \left(Wu(x) + b + \int_{\Omega} K(x, y)u(y)dy \right), \quad (1)$$

where u is the input function, σ is a non-linear activation function, W and b represent a local linear transformation, and $K(x, y)$ denotes a learnable kernel function capturing non-local spatial dependencies. To computationally scale this framework to large three-dimensional configurations, we adopt an anisotropically factorized Fourier neural operator inspired by the FNO and F-FNO architectures [26, 25]. Instead of computing a single fully coupled 3-D spectral kernel, each layer is decomposed into a sequential composition of a 2-D Fourier layer acting on independent xy -slices, a 1-D Fourier layer along the axial (z) direction, and a pointwise channel-mixing branch.

More explicitly, let $v^{(\ell)}$ denote the hidden feature at layer ℓ . The lateral (xy) spectral branch applies a two-dimensional Fourier operator independently on each z -slice:

$$u_o^{(xy, \ell)}(x, y, z) = \mathcal{F}_2^{-1} \left[\sum_{i=1}^{d_v} \mathcal{K}_{oi}^{xy, (\ell)}(k_x, k_y) \mathcal{F}_2(v_i^{(\ell)}(\cdot, \cdot, z)) \right]. \quad (2)$$

The lateral kernel $\mathcal{K}_{oi}^{xy, (\ell)}(k_x, k_y)$ is nonzero only for $|k_x| \leq k_x^{\max}$ and $|k_y| \leq k_y^{\max}$, corresponding to a hard spectral truncation at the retained Fourier modes.

The axial (z) branch is then applied to the laterally filtered feature:

$$u_o^{(z, \ell)}(x, y, z) = \mathcal{F}_1^{-1} \left[\sum_{i=1}^{d_v} \mathcal{K}_{oi}^{z, (\ell)}(k_z) \mathcal{F}_1(u_i^{(xy, \ell)}(x, y, \cdot)) \right], \quad (3)$$

where $\mathcal{K}_{oi}^{z, (\ell)}(k_z)$ is nonzero only for $|k_z| \leq k_z^{\max}$.

Finally, these filtered features are combined and updated via a local pointwise multilayer perceptron (MLP) with a residual connection:

$$v_o^{(\ell+1)}(\mathbf{x}) = v_o^{(\ell)}(\mathbf{x}) + \sigma \left(\sum_{j=1}^{d_h} W_{2,oj}^{(\ell)} \sigma \left(\sum_{i=1}^{d_v} W_{1,ji}^{(\ell)} \left(u_i^{(xy, \ell)}(\mathbf{x}) + u_i^{(z, \ell)}(\mathbf{x}) \right) + b_{1,j}^{(\ell)} \right) + b_{2,o}^{(\ell)} \right). \quad (4)$$

Here, σ denotes the GeLU activation. GeLU is also applied after the pointwise MLP in all intermediate layers, whereas this outer activation is omitted in the final layer. The matrices $W_1^{(\ell)}$ and $W_2^{(\ell)}$, together with the bias vectors $b_1^{(\ell)}$ and $b_2^{(\ell)}$, are the learnable parameters of the pointwise MLP at layer ℓ . The indices i, j , and o denote the input, hidden, and output channels, respectively. Here, d_v is the number of feature channels in the Fourier layer and d_h is the hidden-channel dimension of the MLP.

By composing L such operator layers, the network iteratively extracts and refines the non-local spatial interactions inherent to the electromagnetic scattering process. To initialize the forward evaluation, the mask geometry is encoded by a binary material-indicator field $a_{\text{bin}}(\mathbf{x}) \in \{0, 1\}$. Before network evaluation, this indicator is Gaussian-smoothed on the computational grid to produce a regularized input field $a(\mathbf{x}) \in [0, 1]$. The regularized input field is then fed into the PINO as its structural input. This input field, concatenated with the grid coordinates (x, y, z) as additional channels, is first lifted to a higher-dimensional channel space via a shallow pointwise

linear transformation, commonly referred to as a lifting layer, to formulate the initial hidden state $v^{(0)}(\mathbf{x})$. Following the sequential application of the L factorized Fourier layers, the final hidden representation $v^{(L)}(\mathbf{x})$ is projected back to the physical target space through a local decoding layer.

Consequently, the complete end-to-end architecture acts as a parameterized surrogate operator $\mathcal{G}_\theta$, where θ denotes the set of all learnable weights and biases. Ultimately, the network maps the regularized input field to the predicted 3-D vectorial scattered electric field, expressed as

$$\mathcal{G}_\theta : a(\mathbf{x}) \mapsto \hat{\mathbf{E}}_{\text{scat}}(\mathbf{x}) \in \mathbb{C}^3. \quad (5)$$

The optimal parameters θ are subsequently determined by minimizing a designated loss function, thereby training the network to accurately approximate the true solution operator.

2.2 Loss function for physics-informed neural operator

The loss function of the PINO surrogate model is constructed using the PSFD equation with the background decomposition method [24]. The detailed formulation can be found in our previous paper; here, we briefly summarize the key points.

At every training iteration, the model predicts the scattered field $\hat{\mathbf{E}}_{\theta, \text{scat}}$ strictly within the mask scatterer domain. Given this prediction, the reflected background field $\mathbf{E}_{\text{ML}}(\mathbf{r})$ is reconstructed from the total field at the computationally defined multilayered interface $z = z_{\text{ml}}$. Specifically, we define a reflection operator $\mathcal{R}_{\text{ml}}$ based on the transfer-matrix method (TMM) that maps the transverse field components at the interface $z = z_{\text{ml}}$ to the reflected background field, expressed as the functional mapping:

$$\mathcal{R}_{\text{ml}} : \left[\hat{\mathbf{E}}_{\theta, \text{scat}}(\mathbf{r}) + \mathbf{E}_{\text{inc}}(\mathbf{r}) \right]_{\parallel, z=z_{\text{ml}}} \mapsto \mathbf{E}_{\text{ML}}(\mathbf{r}). \quad (6)$$

This operator inherently incorporates (i) upgoing/downgoing field decomposition, (ii) TE/TM mode decomposition, (iii) the TMM-based reflection response at the fictitious interface, and (iv) an analytical phase-matching propagation to restore the full 3D spatial field distribution $\mathbf{E}_{\text{ML}}(\mathbf{r})$. This reconstructed field is added to the incident field $\mathbf{E}_{\text{inc}}$ to update the total background field. By self-consistently coupling the fields in the mask-only domain to the analytically evaluated multilayer response at every iteration, the formulation retains the full-vector Maxwell coupling between the mask and the planar multilayer without invoking a finite-order Born approximation. Remaining errors arise from finite spatial discretization, PML truncation, interface regularization, iterative solver tolerances, and neural-operator approximation.

The physics-informed loss function $\mathcal{J}(\theta)$ is formulated by defining an effective source term, representing the total background fields including the incident field, as:

$$\mathbf{b}_{\text{bg}}(\mathbf{r}) = k_0^2 \Delta \varepsilon(\mathbf{r}) \left(\mathbf{E}_{\text{inc}}(\mathbf{r}) + \mathbf{E}_{\text{ML}}(\mathbf{r}) \right). \quad (7)$$

Here, $\Delta \varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r}) - \varepsilon_{\text{bg}}(\mathbf{r})$.

Consequently, the mean-squared residual of the governing vectorial PSFD equation is minimized together with a frequency-weighted penalty that emphasizes higher spatial-frequency components:

$$\begin{aligned} \mathcal{J}(\theta) = & \frac{1}{N} \sum_{\alpha \in \{x, y, z\}} \left\| \left[\mathcal{L} \hat{\mathbf{E}}_{\theta, \text{scat}} - \mathbf{b}_{\text{bg}} \right]_{\alpha} \right\|_2^2 \\ & + \frac{\lambda_{\text{hf}}}{N} \sum_{\alpha \in \{x, y, z\}} \left\| w(\nu) \mathcal{F}_3 \left[\left(\mathcal{L} \hat{\mathbf{E}}_{\theta, \text{scat}} - \mathbf{b}_{\text{bg}} \right)_{\alpha} \right] \right\|_2^2, \end{aligned} \quad (8)$$

where $N = N_x N_y N_z$ is the number of spatial grid points, α denotes the Cartesian field component, $\mathcal{F}_3$ is the three-dimensional discrete Fourier transform, and

$$w(\boldsymbol{\nu}) = w_0 + \eta_{\text{hf}} |\boldsymbol{\nu}|, \quad |\boldsymbol{\nu}| = \sqrt{\nu_x^2 + \nu_y^2 + \nu_z^2}. \quad (9)$$

is a frequency-dependent weight that increasingly emphasizes residual components at higher spatial frequencies. Here, λ_{hf} controls the contribution of the frequency-weighted penalty, w_0 is the baseline weight, and η_{hf} determines the strength of the high-frequency emphasis. The operator $\mathcal{L}$ is the vectorial Helmholtz operator given by

$$\begin{aligned} \mathcal{L} &= \nabla \times \nabla \times - k_0^2 \varepsilon(\mathbf{r}) \mathbf{I} \\ &= \begin{bmatrix} -\partial_y \partial_y - \partial_z \partial_z - k_0^2 \varepsilon & \partial_x \partial_y & \partial_x \partial_z \\ \partial_y \partial_x & -\partial_x \partial_x - \partial_z \partial_z - k_0^2 \varepsilon & \partial_y \partial_z \\ \partial_z \partial_x & \partial_z \partial_y & -\partial_x \partial_x - \partial_y \partial_y - k_0^2 \varepsilon \end{bmatrix}. \end{aligned} \quad (10)$$

Here, ∂_α denotes differentiation with respect to the coordinate α . In the numerical implementation, all derivatives, including the mixed terms, are evaluated pseudo-spectrally with PML coordinate stretching as detailed in Ref. [24]. Precomputed simulation results are not used for training the PINO surrogate model.

3 Results

In this section, we present numerical results demonstrating the accuracy of the background-decomposed PSFD formulation, the inference accuracy of the trained PINO surrogate, its scalability to simulation domains at resolutions finer than those used during training, and the convergence acceleration achieved through PINO warm-start initialization. Throughout all simulations, the mask structure is illuminated by a TE-polarized (E_y) plane wave at the EUV wavelength $\lambda = 13.5$ nm under normal incidence. Following Ref. [28], the multilayer substrate is modeled as a Mo/Si distributed Bragg reflector with 40 bilayer pairs, each with a nominal period of 7 nm, for a total stack thickness of approximately 280 nm.

Background-decomposition. We first show the accuracy advantage of the background-decomposed PSFD formulation [24] over a standard full-domain PSFD solver that numerically discretizes the entire multilayer stack. This accuracy advantage carries over naturally to the PINO surrogate, since its physics-informed objective is constructed from the background-decomposed PSFD residual. If the PINO were instead trained using the residual of a full-domain PSFD solver, systematic multilayer discretization errors would become entangled with neural-operator approximation errors.

Figure 1(a) shows the scattered intensity $|E_y^{\text{scat}}|^2$ obtained by the background-decomposed PSFD solver within the computational domain spanning only the 50 nm-thick Ta absorber region (dashed outline), with the 280 nm-thick underlying multilayer stack excluded from explicit discretization. The multilayer-reflected background field, $\mathbf{E}_{\text{ML}}(\mathbf{r})$, is updated at every iteration via TMM and imposed as a self-consistent reflected-background contribution, as defined in Eq. (7), while the iterative PSFD solver resolves only the absorber-induced scattered field $\mathbf{E}^{\text{scat}}$. Removing the multilayer from the computational domain reduces the domain height substantially, leading to a corresponding decrease in both memory overhead and the number of solver iterations. The

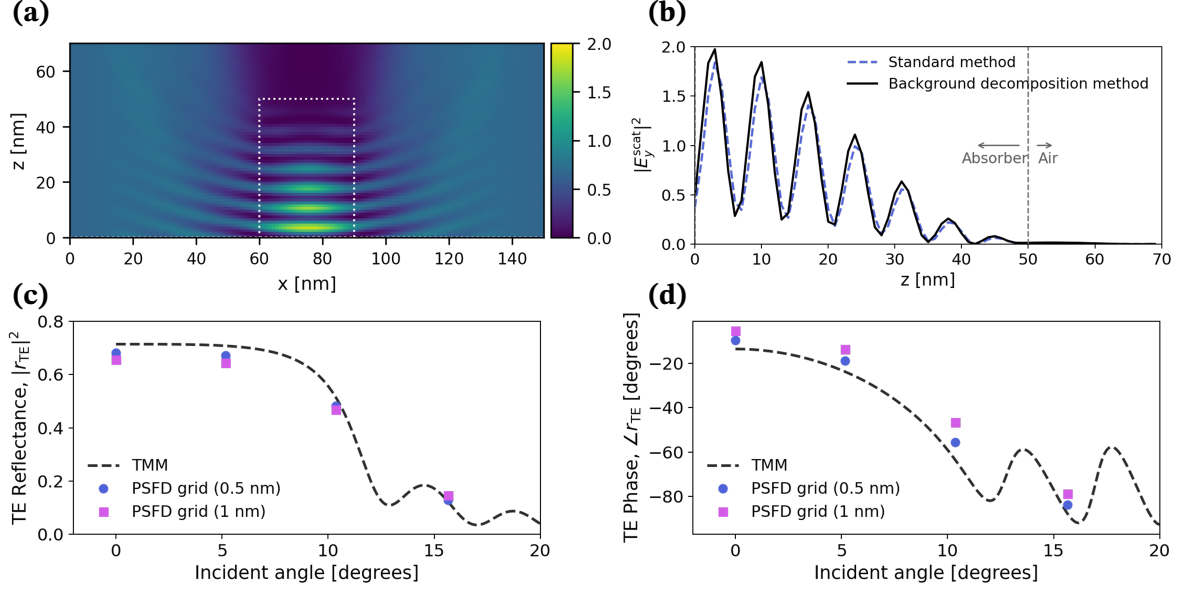


Figure 1: (a) Scattered intensity $|E_y^{\text{scat}}|^2$ computed by the background-decomposed PSFD solver with the computational domain covering only the 50 nm Ta absorber region (dashed outline), (b) Vertical profile of $|E_y^{\text{scat}}|^2$ along the z -axis at the center of the absorber region for both solvers at $\Delta x = 1$ nm, (c, d) TE reflectance $|r_{\text{TE}}|^2$ and reflection phase $\arg(r_{\text{TE}})$ of the bare multilayer computed by the standard solver at $\Delta x = 1$ nm and 0.5 nm, compared with the exact TMM result.

reduction in vertical (z -axis) grid points can be redistributed to enlarge the in-plane (xy -plane) simulation area, directly enabling large-area lateral scaling at a fixed memory budget.

Figure 1(b) compares the line profile of the scattered intensity $|E_y^{\text{scat}}|^2$ extracted at $x = 75$ nm from the 2D field shown in Figure 1(a), as obtained by the background-decomposed PSFD solver (black line) and a standard PSFD solver that explicitly discretizes the entire multilayer stack (dotted blue line), both computed at a uniform grid spacing of $\Delta x = \Delta y = \Delta z = 1$ nm. A measurable discrepancy in $|E_y^{\text{scat}}|^2$ is observed between the two methods along the entire profile. This discrepancy is attributed to a discretization error that arises when numerically resolving the optically thick multilayer. The complex interference between upward- and downward-propagating modes within the stack is sensitive to the grid spacing. A finite grid spacing introduces a grid-size-dependent phase error that causes a fraction of the reflected scattered energy to leak numerically rather than being faithfully returned to the absorber region.

Figure 1(c) and 1(d) show the TE-polarized reflectance $|r_{\text{TE}}|^2$ and the reflection phase $\arg(r_{\text{TE}})$ of the bare multilayer stack without the absorber, as computed by the standard solver at grid spacings $\Delta x = 1$ nm and $\Delta x = 0.5$ nm. Both quantities are compared with the analytically exact TMM results to quantify the numerical leakage introduced by discretizing the multilayer stack. With discrete grid spacing Δx , the transverse wavenumbers supported by the Fourier grid are restricted to $k_{x,m} = 2\pi m / (N_x \Delta x)$, so that only a discrete set of incidence angles $\theta_m = \arcsin(k_{x,m}/k_0)$ can be faithfully represented without inducing angular dispersion artifacts. For the two grid spacings considered here, the representable angles within the range 0° – 20° are $\theta = 0^\circ, 5.16^\circ, 10.37^\circ$, and 15.66° , so the comparison is accordingly restricted to these angles. Both grid spacings yield $|r_{\text{TE}}|^2$ and $\arg(r_{\text{TE}})$ values that systematically deviate from the TMM references, with the deviation decreasing as Δx is reduced. The systematic decrease in the deviation as Δx is reduced is consistent

with a multilayer-discretization error that converges under grid refinement. Achieving this limit requires increasing the number of grid points N_x at a fixed physical domain size, which scales the memory requirement and imposes a prohibitive memory overhead. Thus, the background-decomposed PSFD solver avoids the multilayer-discretization error and memory overhead associated with explicitly meshing the stack.

PINO surrogate. The PINO is trained on a fixed dataset of approximately 16,000 mask designs drawn from the LithoBench benchmark library, with mini-batches sampled randomly at each training iteration. For inference, the trained PINO is evaluated on a separate set of 100 held-out mask designs that were not included in the training dataset. Each sample is defined on a computational domain of $600 \times 600 \times 200$ nm³ at grid spacings of $\Delta x = \Delta y = 4$ nm and $\Delta z = 2$ nm.

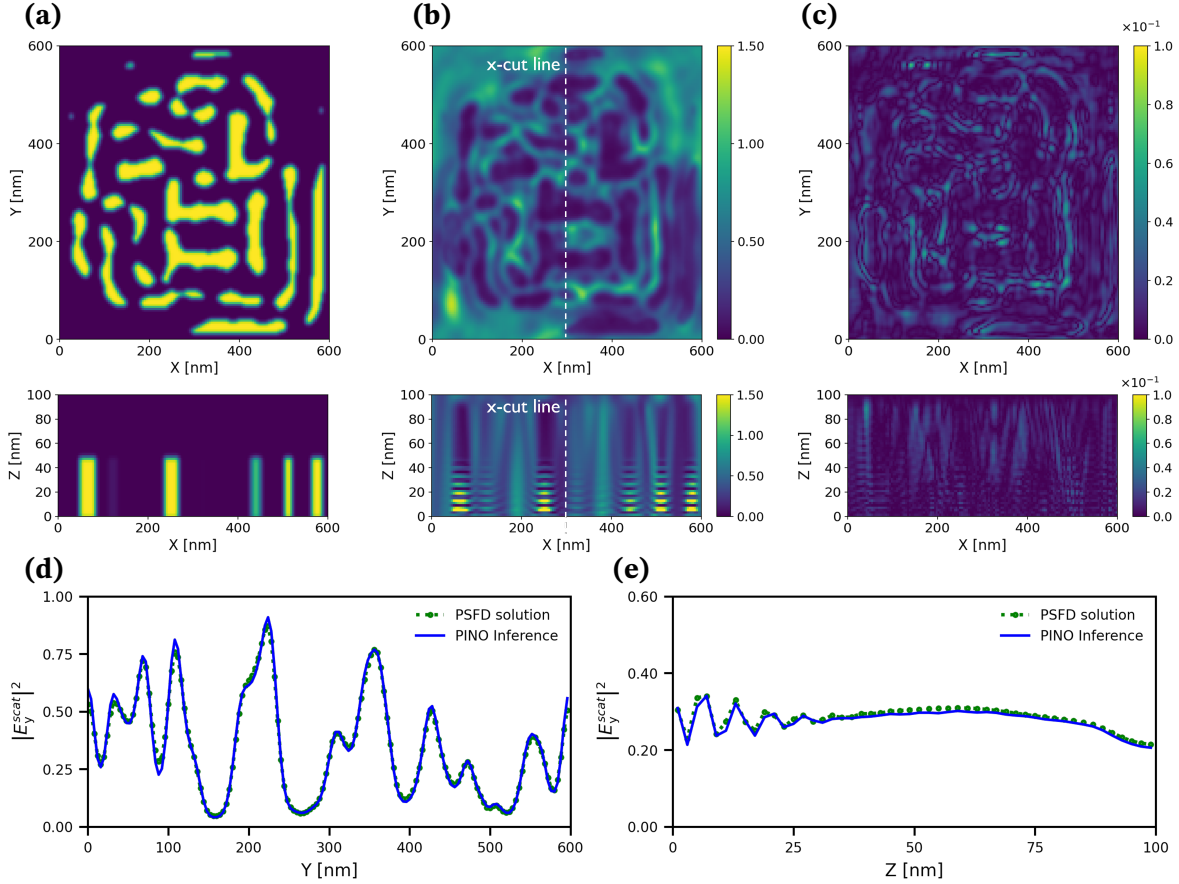


Figure 2: (a) Regularized input field encoding the material distribution of a representative test mask, shown as horizontal (xy -plane) and vertical (xz -plane) cross-sections. (b) Scattered intensity $|E_y^{\text{scat}}|^2$ just above the absorber as inferred by the trained PINO surrogate, shown in the xy - and xz -planes. (c) Mean absolute error of the PINO inference with respect to the background-decomposed PSFD reference, shown in the xy - and xz -planes. (d, e) One-dimensional profiles of $|E_y^{\text{scat}}|^2$ along the y -axis and z -axis at the x -cut indicated in (b), comparing the PINO inference with the background-decomposed PSFD reference.

Figure 2(a) shows horizontal (xy -plane) and vertical (xz -plane) cross-sections of the regularized input field for a representative test mask. Since a discontinuous binary interface contains

spatial-frequency components beyond those representable on a finite Fourier grid, its direct truncation produces Gibbs ringing and can degrade convergence during training with the spectral loss. A Gaussian smoothing is therefore applied to the binary indicator [29, 30, 31] to obtain the regularized input field. The same Gaussian smoothing is applied to the structure used by the rigorous PSFD solver before constructing the material distribution. Figure 2(b) shows the scattered intensity $|E_y^{\text{scat}}|^2$ evaluated just above the absorber, as predicted by the trained PINO surrogate. Figure 2(c) shows the corresponding mean absolute error (MAE). All pointwise errors and MAEs reported in this work are computed from absolute differences between the PINO-predicted and PSFD-reference scattered intensities $|E_y^{\text{scat}}|^2$. The PINO is trained on a single NVIDIA A100 80 GB GPU until convergence, at which point the physics residual reaches approximately 10^{-2} . The correspondence with the background-decomposed PSFD solver is more clearly quantified in Figure 2(d) and 2(e), which compare one-dimensional profiles of $|E_y^{\text{scat}}|^2$ extracted along the y -axis in the xy -plane and along the z -axis in the xz -plane, respectively. Both profiles show excellent agreement between the PINO prediction and the PSFD reference, with an MAE of approximately 7×10^{-3} .

Application to finer inference grids. We next examine the transfer of the trained PINO surrogate to an inference grid finer than that used during training. In this study, the target inference grid is $N_x \times N_y \times N_z = 1200 \times 1200 \times 100$ at $\Delta x = \Delta y = 1$ nm and $\Delta z = 2$ nm, corresponding to the same physical domain of $L_x \times L_y \times L_z = 1200 \times 1200 \times 200$ nm³. As a test case for this evaluation, Figure 3 shows a representative regularized mask design from the test set, displayed as two-dimensional cross-sections in the xy -, xz -, and yz -planes.

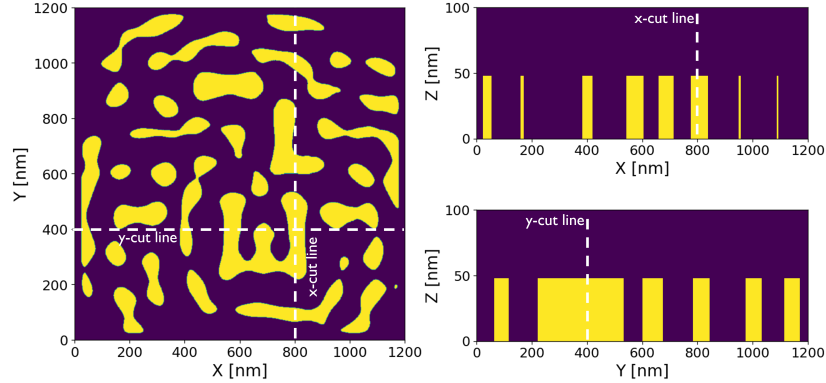


Figure 3: Regularized input field of a representative held-out mask geometry, displayed as two-dimensional cross-sections in the xy -, xz -, and yz -planes.

The dashed lines indicate the cut positions used to extract the xy - and yz -plane cross-sections, and the z -cut is taken at the center of the absorber layer. Direct training of the FNO on the target inference grid of $N_x \times N_y \times N_z = 1200 \times 1200 \times 100$ is computationally challenging, since training requires simultaneously holding in device memory the intermediate layer activations for backpropagation, the gradient tensors, and the Adam optimizer state buffers, all of which scale as $\mathcal{O}(N_x N_y N_z)$. None of these are required during inference [26, 16, 32]. The training bottleneck therefore imposes a practical upper limit on the domain size under given GPU memory constraints. To remain within the GPU memory budget, the PINO is trained on a coarser grid of $N_x \times N_y \times N_z = 300 \times 300 \times 100$ at grid spacings of $\Delta x = \Delta y = 4$ nm and $\Delta z = 2$ nm, correspond-

ing to a physical domain of $L_x \times L_y \times L_z = 1200 \times 1200 \times 200 \text{ nm}^3$. Inference is then performed at finer grid spacings of $\Delta x = \Delta y = 1 \text{ nm}$ and $\Delta z = 2 \text{ nm}$, increasing the grid resolution to $N_x \times N_y \times N_z = 1200 \times 1200 \times 100$.

The scattered intensity $|E_y^{\text{scat}}|^2$ computed for this mask geometry is shown in Figure 4, where Figure 4(a) presents the PINO inference and Figure 4(b) the corresponding background-decomposed PSFD reference at $\Delta x = \Delta y = 1 \text{ nm}$, with their pointwise absolute difference given in Figure 4(c). The quantitative comparison yields an MAE of approximately 3×10^{-2} . These values are roughly half an order of magnitude larger than the corresponding errors observed in Figure 2, where the grid spacings are matched. The errors at the edges of the geometries are more pronounced as seen in Figure 2(c) and Figure 4(c). The difference can be attributed to higher spatial frequencies supported by the finer inference grid but not represented on the coarser training grid.

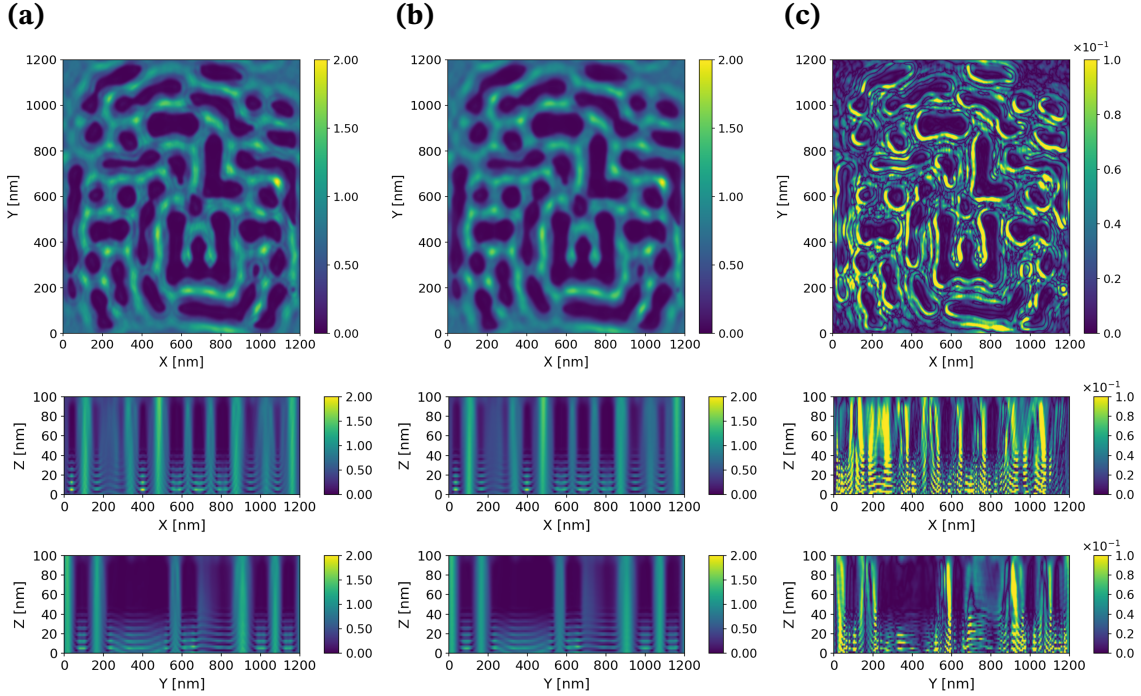


Figure 4: Comparison of the scattered intensity $|E_y^{\text{scat}}|^2$ for the test mask of Fig. 3 at $\Delta x = \Delta y = 1 \text{ nm}$, four times finer than the training grid spacing. (a) background-decomposed PSFD reference, (b) PINO inference, and (c) pointwise absolute difference between (a) and (b).

PINO warm-start for preconditioned PSFD with Krylov iteration. We next investigate the trained PINO model as a warm-start initializer for the preconditioned Krylov solver underlying the iterative background-decomposed PSFD scheme. This approach is designed to improve the early-stage convergence behavior and reduce the number of iterations to meet practical convergence criteria. The linear system arising from the background-decomposed PSFD discretization is solved using the BiCGSTAB algorithm [33, 34], which is a Krylov subspace method well-suited to complex, non-Hermitian systems of the form encountered in frequency-domain electromagnetic scattering.

Right preconditioning [35, 33] introduces an auxiliary variable for the Krylov iteration. With this approach, convergence can be improved by transforming the original system into a more fa-

favorable spectral form through a preconditioner that approximates selected features of the inverse system matrix $\mathbf{A}^{-1}$. In addition, initialization [33] supplies a starting iterate $\mathbf{E}^{(0)}$ that reduces the initial residual and shortens the early transient correction phase before the asymptotic regime is reached. In this work, three configurations are compared: (i) null preconditioner and null initialization, (ii) the spectral damping preconditioner [24] and null initialization, and (iii) the spectral damping preconditioner combined with PINO-based warm-start initialization.

The discretized PSFD system is written as $\mathbf{A}\mathbf{E} = \mathbf{b}$. Convergence is then assessed using the relative residual, which provides a dimensionless, amplitude-invariant measure:

$$r_{\text{rel}}^{(k)} = \frac{\|\mathbf{b} - \mathbf{A}\mathbf{E}^{(k)}\|_2}{\|\mathbf{b}\|_2}. \quad (11)$$

Figure 5(a) plots $r_{\text{rel}}^{(k)}$ as a function of iterations (k) for the three configurations. In the null case, the relative residual decreases slowly with increasing iteration count, reaching approximately 5×10^{-2} after 1,000 iterations. The scattered field progresses from a zero initial state toward an approximate solution over the course of 1,000 iterations. At this point, the dominant spectral components of the field have been largely recovered, as seen in Figure 6(a). As shown later, this residual level represents a practically relevant operating point for rapid field computation, since the resulting scattered field shows reasonable accuracy for estimating variations in lithographically relevant feature dimensions even before fully converging.

Applying the scalar spectral-damping right preconditioner substantially steepens the observed residual decay, as shown in Figure 5(a). The spectral damping operator $\mathcal{M}$ is applied independently to each field component through a scalar Fourier multiplier:

$$\mathcal{F}_3[\mathcal{M}\mathbf{x}](\mathbf{k}) = m(\mathbf{k}) \mathcal{F}_3[\mathbf{x}](\mathbf{k}), \quad (12)$$

where the three-dimensional Fourier transform is applied componentwise and

$$m(\mathbf{k}) = m_{\min} + (1 - m_{\min}) \begin{cases} 1, & |\mathbf{k}| \leq k_{\text{ref}}, \\ \left(\frac{k_{\text{ref}}^2}{|\mathbf{k}|^2 + \epsilon} \right)^\gamma, & |\mathbf{k}| > k_{\text{ref}}. \end{cases} \quad |\mathbf{k}|^2 = k_x^2 + k_y^2 + k_z^2. \quad (13)$$

Here, $m_{\min}$ is the multiplier floor, k_{ref} is the cutoff wavenumber, γ controls the high-frequency decay, and ϵ is a small numerical regularizer. With $m_{\min} = 0$, $\gamma = 1$, $\epsilon \rightarrow 0$, and $k_{\text{ref}} = k_{\text{bg}}$, Eq. (13) reduces to the isotropic spectral damping baseline $D(k) = \min(1, k_{\text{bg}}^2/k^2)$ used in Ref. [24]. The present form extends that baseline by introducing a nonzero floor and a tunable decay exponent. With this approach, the Krylov solver solves for an auxiliary variable $\mathbf{x}$, from which the physical field $\mathbf{E}$ is recovered as

$$\mathbf{A}\mathbf{E} = \mathbf{A}\mathcal{M}\mathbf{x} = \mathbf{b} \quad (14)$$

This makes the Krylov iteration more efficient, as the preconditioner $\mathcal{M}$ effectively conditions high-frequency components that are typically associated with slow convergence. The PINO warm start extends this formulation by introducing a nonzero base field through an affine right-correction formulation. Letting $\mathbf{E}_{\text{base}}$ denote the initial field, the solution is written as

$$\mathbf{E} = \mathbf{E}_{\text{base}} + \mathcal{M}\mathbf{y}, \quad (15)$$

and the Krylov solver is applied to

$$\mathbf{A}\mathcal{M}\mathbf{y} = \mathbf{b} - \mathbf{A}\mathbf{E}_{\text{base}}. \quad (16)$$

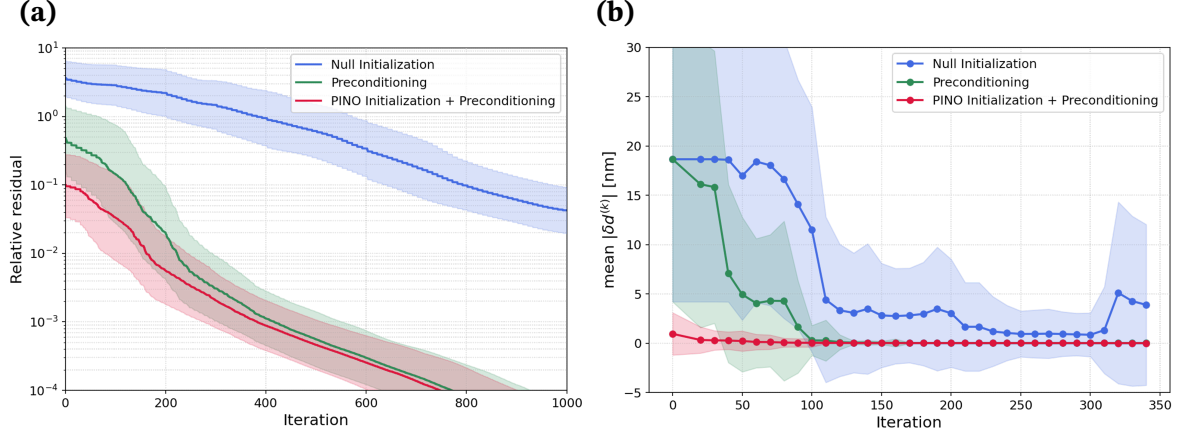


Figure 5: (a) Relative residual as a function of iterations for the three solver configurations. (b) Mean and standard deviation of the wafer-equivalent near-field threshold-width deviation $|\delta d^{(k)}|$ over iterations for the three configurations.

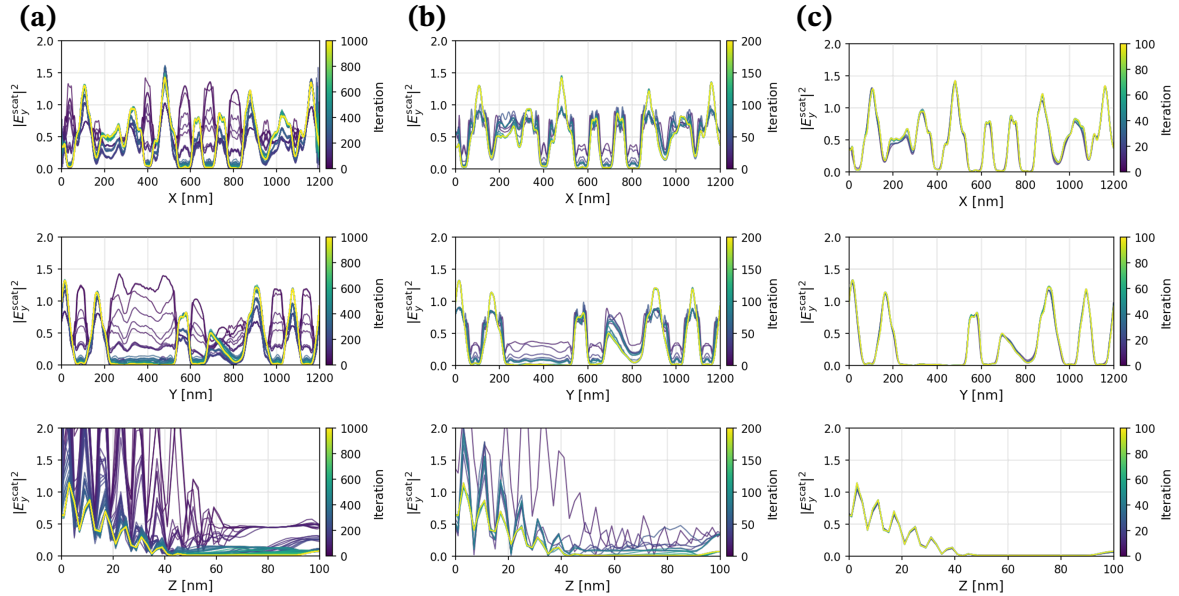


Figure 6: Convergence trends of the background-decomposed PSFD solver under three configurations, evaluated by the iteration-to-iteration change in the reflected scattered field $\mathbf{E}^{\text{scat}}$ at the top surface of the absorber, along the x-, y- and z-axis indicated in Figure 3. (a) Null initialization without preconditioning, shown up to 1,000 iterations. (b) Scalar spectral-damping right preconditioning with null initialization, shown up to 200 iterations. (c) Scalar spectral-damping right preconditioning with PINO warm-start initialization, shown up to 100 iterations.

For null initialization, $\mathbf{E}_{\text{base}} = \mathbf{0}$ and $\mathbf{y} = \mathbf{x}$, so the affine correction system reduces to the standard right-preconditioned formulation in Eq. (14). The warm-start field $\mathbf{E}_{\text{base}}$ is therefore preserved, while the damping operator is applied only to the iterative correction $\mathbf{y}$. The scalar multiplier leaves low-spatial-frequency corrections unchanged and attenuates higher-frequency corrections down to the floor m_{min} . This modification empirically improves the convergence behavior for the problems considered here. At a relative residual of $r_{\text{rel}}^{(k)} \approx 5 \times 10^{-2}$, where the MAE falls below 1×10^{-3} , the preconditioned-only configuration reaches this accuracy level in approximately 150 iterations, whereas the null configuration requires approximately 1,000 iterations.

When the PINO warm-start is additionally applied, the iteration begins from an approximately one-order-of-magnitude lower initial residual, shifting the entire decay curve downward. The same relative-residual level of 5×10^{-2} is reached in approximately 75 iterations, corresponding to a factor-of-2 reduction relative to the preconditioned-only case and a factor-of-13 reduction relative to the null case. The representative end-to-end wall-clock times are approximately 30 minutes, 4 minutes, and 2 minutes for the null, preconditioned-only, and PINO-warm-start configurations, respectively. For the PINO case, the reported time includes model loading, neural-operator inference, grid interpolation, and field conversion. As shown in Figure 6(c), the scattered field starts close to the final solution and changes only marginally over the course of the iteration, with the remaining updates confined to fine-scale refinements. This early transient phase is of primary practical interest, as the dominant field components are established here and practical convergence criteria may already be satisfied.

The preconditioned null-initialized curve exhibits a slightly steeper decay during part of the early transient than the PINO-initialized curve. This difference reflects the different spectral content and magnitudes of their initial residuals. The PINO warm start has already removed a substantial portion of the dominant initial field error, whereas the null-initialized solve begins with a larger residual distributed over a broader range of spectral components. As the iteration proceeds toward tighter tolerances, the two preconditioned configurations exhibit increasingly similar empirical decay rates. Accordingly, the benefit of PINO initialization becomes progressively less pronounced.

Practical assessment of scattered-field convergence. Following the threshold-contour procedure used for CD extraction from aerial or resist images, we apply an analogous contour-based metric directly to the scattered-field intensity to assess a practically useful accuracy level without performing full projection-optics or resist simulation. For this, feature widths are extracted and evaluated from crossings of a fixed intensity threshold along horizontal and vertical scan lines. For each segment bounded by two threshold crossings, we track the deviation $\delta d^{(k)} = d_{\text{ref}} - d^{(k)}$, where $d^{(k)}$ is the threshold-contour width at iteration k . The reference value d_{ref} is obtained from the fully converged PSFD solution at $r_{\text{rel}}^{(k)} \approx 10^{-7}$. This metric additionally quantifies the stability of the scattered near-field geometry as the iterative solution converges.

For comparison with wafer-scale lithographic dimensions, the measured contour displacements are divided by four, corresponding to the ideal isotropic 4 $\times$ demagnification of a 0.33-NA EUV optical system. The metric does not include projection-optics filtering, partial coherence, aerial-image formation, or resist-process effects, and should therefore not be directly interpreted as a predicted printed-feature CD error. The intensity field is mapped through the logistic sigmoid

function

$$S(I) = \frac{1}{1 + \exp\left(-\frac{I - I_{\text{th}}}{w}\right)}, \quad (17)$$

where $I = |E_y^{\text{scat}}|^2$, and $I_{\text{th}} = 0.15$ and $w = 0.05$ are the threshold and transition-width parameters adopted in this study.

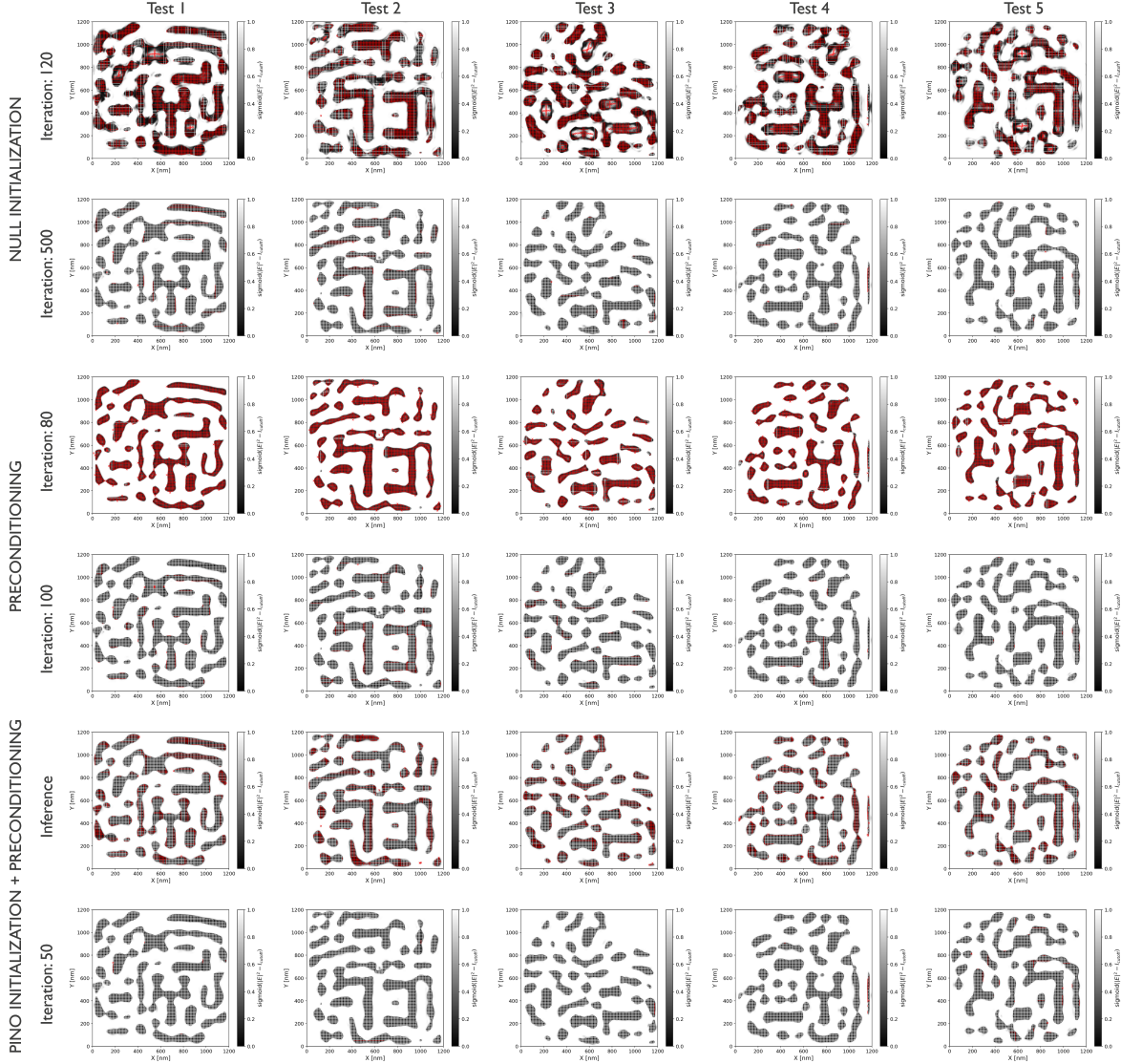


Figure 7: Spatial maps of the wafer-equivalent near-field width error $|\delta d|$ overlaid on the thresholded scattered-intensity field $S(|E_y^{\text{scat}}|^2)$ for the three solver configurations at representative early and late iterations. Red segments indicate scan-line widths with $|\delta d| > 1$ nm, red crosses indicate locations where no resolvable threshold-bounded segment is formed, and white segments indicate $|\delta d| \leq 1$ nm.

The threshold-contour width $d^{(k)}$ is extracted at the $S = 0.5$ contour, which corresponds to $I = I_{\text{th}}$. At each iteration, $d^{(k)}$ is measured along 100 horizontal and 100 vertical scan lines per

pattern and averaged over 100 test patterns not used during training. Figure 5(b) shows the mean and standard deviation of $|\delta d^{(k)}|$ over iterations for the three configurations. In the null case, $|\delta d^{(k)}|$ is initially large and decreases only slowly. Figure 7 visualizes these trends by overlaying scan-line segments on the sigmoid-binarized intensity field at representative iteration snapshots. Scan-line positions where $|\delta d| > 1$ nm are shown in red, whereas those where $|\delta d| \leq 1$ nm are shown in white. For the null configuration, above-threshold segments dominate the map at iteration 120, reflecting widespread above-threshold deviations, while by iteration 500 the fraction of above-threshold segments decreases significantly, confirming the slow but gradual approach to a sub-nanometer near-field threshold-width deviation in wafer-equivalent units.

The spectral damping preconditioner accelerates this decay markedly, reducing both the mean and standard deviation of the near-field threshold-width proxy below 1 nm in wafer-equivalent units within about 150 iterations. The fraction of above-threshold segments in Figure 7 is dramatically reduced from iteration 80 to 100, confirming the observation from the relative residual that the preconditioner accelerates convergence in the early transient correction phase.

With additional PINO warm-start initialization, $|\delta d^{(k)}|$ is already relatively small from the inference stage and the fraction of above-threshold segments is correspondingly low. This method reaches the same sub-nanometer near-field proxy level within about 75 iterations. These results demonstrate that the PINO surrogate can serve as a warm-start initializer for the background-decomposed PSFD solver, reducing the number of iterations required to satisfy the prescribed practical convergence criterion. When strict convergence to $r_{\text{rel}}^{(k)}$ of 10^{-7} is required, however, the required iterations increase dramatically, with the null case requiring about 11,000 iterations, and either preconditioned configuration requiring about 2,000 iterations.

4 Conclusion

We have presented a background-decomposed PSFD framework combined with a PINO surrogate for scalable EUV mask simulation. In this framework, the computational domain is substantially reduced to the mask absorber region, while the multilayer response is incorporated analytically through a TMM-based reflection operator during both training and rigorous simulation.

At its native training resolution, the first PINO model achieves an MAE of approximately 7×10^{-3} on held-out mask geometries. For the finer-grid evaluation, a separately trained PINO model is applied without retraining to a four-times finer lateral discretization of the same physical domain, yielding an MAE of approximately 3×10^{-2} . In this case, the error is largely concentrated near material interfaces, where the finer inference grid resolves spatial-frequency content beyond that represented on the coarser training grid.

The scalar spectral-damping right preconditioner reduces the iteration count by a factor of approximately 6.7 relative to the null case under the prescribed convergence criterion. When combined with the preconditioner, PINO warm-start initialization accelerates the early iteration phase, reducing the number of iterations required to reach a practically useful accuracy level by a factor of approximately 13 relative to the null baseline.

The corresponding representative end-to-end wall-clock times are about 30 minutes for null initialization and about 1–2 minutes for PINO warm-start initialization, with the latter including model loading, inference, interpolation, and field conversion. When tight convergence is required, the two preconditioned configurations reach the target with similar iteration counts, and the benefit associated with PINO initialization becomes less pronounced. The proposed framework provides an effective route to rapid large-scale field computation by combining fast PINO-

based initialization and preconditioning with the accuracy of the rigorous PSFD solver.

5 Acknowledgments

The authors express gratitude to Seungjin Lee, Pervaiz Kareem, Hazem Mesilhy and Xuelong Shi for valuable discussions during the initial stages of this work. This work has been enabled in part by the NanoIC pilot line. The acquisition and operation are jointly funded by the Chips Joint Undertaking, through the European Union's Digital Europe (101183266) and Horizon Europe programs (101183277), as well as by the participating states Belgium (Flanders), France, Germany, Finland, Ireland and Romania. For more information, visit nanoic-project.eu.

References

- [1] A. Taflove, *Advances in Computational Electrodynamics: The Finite-difference Time-domain Method*, Artech House (1998).
- [2] W.C. Chew, *Waves and Fields in Inhomogeneous Media*, Wiley-IEEE Press, New York (1995).
- [3] K. Yee, *Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media*, *IEEE Trans. Antennas Propagat.* **14** (1966) 302.
- [4] M.G. Moharam, E.B. Grann, D.A. Pommet and T.K. Gaylord, *Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings*, *Journal of the Optical Society of America A* **12** (1995) 1068.
- [5] H.H. Hopkins, *On the diffraction theory of optical images*, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **217** (1953) 408.
- [6] A.K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, WA (2001).
- [7] T.V. Pistor, K. Adam and A. Neureuther, *Rigorous simulation of mask corner effects in extreme ultraviolet lithography*, *J. Vac. Sci. Technol. B* **16** (1998) 3449.
- [8] A. Erdmann and P. Evanschitzky, *Modeling of mask effects in optical and euv lithography*, in *Photomask and Next-Generation Lithography Mask Technology XIII*, vol. 6283, p. 62831B, SPIE, 2006.
- [9] A. Erdmann, *Mask 3d effects in euv lithography*, in *Emerging Lithographic Technologies IX*, vol. 5751, pp. 553–564, SPIE, 2005.
- [10] T.V. Pistor, *Electromagnetic simulation of extreme ultraviolet lithography masks*, in *Emerging Lithographic Technologies VI*, vol. 4688, pp. 1031–1041, SPIE, 2002.
- [11] P. Liu, Y. Cao, L. Chen and J. Ye, “Three-dimensional mask model for photolithography simulation.” US Patent 7703069B1, 2010.

- [12] K. Adam and A.R. Neureuther, *Domain decomposition method for rapid calculation of photomask near fields*, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* **20** (2002) 2585.
- [13] T. Watanabe, H. Suzuki, M. Takahashi, T. Matsunawa and S. Nojima, *Fast mask near-field calculation method using deep learning*, *Journal of Micro/Nanolithography, MEMS, and MOEMS* **18** (2019) 044001.
- [14] N. Jia and E.Y. Lam, *Fast aerial image calculation for extreme ultraviolet lithography under partly coherent illumination*, *Journal of Vacuum Science & Technology B* **32** (2014) 021601.
- [15] M. Raissi, P. Perdikaris and G.E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, *Journal of Computational Physics* **378** (2019) 686.
- [16] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu et al., *Physics-Informed Neural Operator for Learning Partial Differential Equations*, 2023. 10.48550/arXiv.2111.03794.
- [17] A. Erdmann, C. Schwemmer and P. Evanschitzky, *Physics-informed neural networks for computational lithography modeling and optimization*, *Journal of Micro/Nanopatterning, Materials, and Metrology* **23** (2024) 011005.
- [18] V. Medvedev, A. Erdmann and A. Roskopf, *3D mask simulation and lithographic imaging using physics-informed neural networks*, in *Optical and EUV Nanolithography XXXVII*, SPIE, 2024, DOI.
- [19] V.A. Es'kin and E.V. Ivanov, *Physics-informed neural networks and neural operators for a study of EUV electromagnetic wave diffraction from a lithography mask*, in *2025 Days on Diffraction (DD)*, 2025, DOI.
- [20] J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Garcia, M. Tegmark et al., *Nanophotonic particle simulation and inverse design using artificial neural networks*, *Science Advances* **4** (2018) eaar4206.
- [21] Y. Khoo, J. Lu and L. Ying, *Solving parametric scattering problems via neural operators*, *Journal of Computational Physics* **443** (2021) 110542.
- [22] P. Evanschitzky and A. Erdmann, *Three dimensional EUV simulations: a new mask near field and imaging simulation system*, in *Emerging Lithographic Technologies IX*, SPIE, 2005, DOI.
- [23] J. Zhang, Y. Li and L. Wang, *Fast rigorous mask model for extreme ultraviolet lithography*, *Applied Optics* (2020) .
- [24] S. Lee, W. Gillijns and D. Kim, *Pseudo-spectral frequency-domain method with background field decomposition and Green's function preconditioner for electromagnetic scattering problem in EUV lithography*, 2026. 10.48550/arXiv.2606.25541.
- [25] A. Tran, A. Mathews, L. Xie and C.S. Ong, *Factorized fourier neural operators*, in *International Conference on Learning Representations*, 2023 [2111 . 13802].

- [26] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart et al., *Fourier Neural Operator for Parametric Partial Differential Equations*, 2021. 10.48550/arXiv.2010.08895.
- [27] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart et al., *Neural Operator: Learning Maps Between Function Spaces*, 2024. 10.5555/3648699.3648788.
- [28] I.A. Makhotkin, M. Wu, V. Soltwisch, F. Scholze and V. Philipsen, *Refined extreme ultraviolet mask stack model*, *J. Opt. Soc. Am. A* **38** (2021) 498.
- [29] B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press (1998).
- [30] Q.H. Liu, *The PSTD algorithm: A time-domain method requiring only two cells per wavelength*, *Microwave and Optical Technology Letters* **15** (1997) 158.
- [31] A. Farjadpour, D. Roundy, A. Rodriguez, M. Ibanescu, P. Bermel, J.D. Joannopoulos et al., *Improving accuracy by subpixel smoothing in the finite-difference time domain*, *Optics Letters* **31** (2006) 2972.
- [32] Z. Li, N. Kovachki, C. Choy, B. Li, J. Kossaifi, S.P. Otta et al., *Geometry-Informed Neural Operator for Large-Scale 3D PDEs*, 2023. 10.48550/arXiv.2309.00583.
- [33] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2nd ed. (2003), 10.1137/1.9780898718003.
- [34] H.A. van der Vorst, *Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems*, *SIAM J. Sci. Stat. Comput.* **13** (1992) 631.
- [35] M. Benzi, *Preconditioning techniques for large linear systems: A survey*, *Journal of Computational Physics* **182** (2002) 418.