Optics & Laser Technology 189 (2025) 113022

Contents lists available at ScienceDirect

Optics and Laser Technology

journal homepage: www.elsevier.com/locate/optlastec

Full length article

Efficient resist modeling and calibration using a Wiener–Padé formulation and convex optimizations

Chunxiao Mu [a], Zhiyang Song [a], Lei Cheng [a], Shaopeng Guo [a], Ke Li [b], Song Zhang [b], Hao Jiang [a,c,*], David H. Wei [b,**], Yanlong Sun [d,e], Jinlong Zhu [a], Shiyuan Liu [a,c,*]

[a] School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, Hubei, China
[b] Yuwei Optica Co., Ltd, Wuhan, 430070, Hubei, China
[c] Optics Valley Laboratory, Wuhan, 430074, Hubei, China
[d] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, 230088, Anhui, China
[e] National Key Laboratory of Human–Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University, Xi'an, 710049, Shanxi, China

ARTICLE INFO

Keywords:
OPC
Computational lithography
Resist modeling
Resist calibration

ABSTRACT

To meet the semiconductor industry's demand for full-chip simulations and optimizations, fast and accurate compact models have become the most widely adopted resist models in Optical Proximity Correction (OPC) for mask synthesis. Current compact models usually rely on adding more model terms within a linear combination framework to handle specific resist characteristics, which increase the computational complexity and decrease the model generality. In this paper, we break away from the conventional linear combination framework and develop a fast, accurate, and versatile Wiener–Padé model. Additionally, we introduce a two-stage calibration strategy based on quadratic convex optimization with effective constraints, in order to improve the efficiency of the calibration process and effectively prevent overfitting and ill-conditioned problem in resist models. Experimental results demonstrate the efficiency and accuracy of our Wiener–Padé models and method of two-stage calibration, while also verifying the benefits of incorporating constraints in the calibration process. Our methodology of Wiener–Padé modeling and quadratic convex optimizations provides a promising solution for advanced resist modeling and calibration in computational lithography.

1. Introduction

The development of semiconductor technology pushes the limits of photolithography to print tiny features on the wafer. As critical dimension (CD) shrinks and feature complexity grows, Optical Proximity Correction (OPC) for mask synthesis has become an indispensable technology for resolution enhancement. OPC is a computational technique used to modify the mask patterns to compensate for the distortions that arise during the lithographic process, ensuring that the printed features match the intended designs. As a key component of OPC tool, the resist model needs capture the nonlinear physical and chemical interactions – such as acid-base neutralization and diffusion reaction – occurring during exposure and development. To meet the demands of full-chip OPC, resist modeling techniques need to be general, fast, and accurate [1,2]. The threshold-based models, such as the constant threshold model and the variable threshold model [3–5], are first proposed. These models correlate the threshold with the spatial information or properties like light intensity or gradient to extract the resist contour from the aerial image. However, these models are limited to some specific lithography conditions. Therefore, more advanced compact models have been introduced and widely adopted as the standard model for current OPC.

Current compact models use a series of operations on aerial images to describe highly nonlinear physical and chemical effects during the resist processes. The most representative one is the CM1, proposed by Yuri Granik and colleagues [6–9]. In this model, neutralization reactions during post-exposure bake can be described as the additive or subtractive transformations applied to the aerial image intensity. Gradient and Laplacian calculations highlight the local slope and curvature of aerial images. Power and root operations mimic nonlinear intensity behaviors such as minimum and maximum response levels. Convolutions with Gauss–Laguerre kernels simulate the diffusion processes during post-exposure bake. By linearly combining these terms, CM1 captures the complex resist effects while maintaining computational efficiency, making it widely applicable in most full-chip OPC lithographic resist modeling tasks. At more advanced nodes, resist characteristics and their underlying physical and chemical mechanisms

* Corresponding authors at: School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, Hubei, China.
** Corresponding author at: Yuwei Optica Co., Ltd, Wuhan, 430070, Hubei, China.
E-mail addresses: hjiang@hust.edu.cn (H. Jiang), david.hq.wei@gmail.com (D.H. Wei), shyliu@hust.edu.cn (S. Liu).

https://doi.org/10.1016/j.optlastec.2025.113022
Received 13 January 2025; Received in revised form 12 March 2025; Accepted 19 April 2025
Available online 10 May 2025

become increasingly complex. To enhance model accuracy and adapt to new challenges, various approaches have been employed to improve the compact models. By introducing new modeling terms, such as adding resist horizontal shrink term and horizontal development bias term, the compact model can be enabled to capture the unique physical effects of NTD resists [10]. And by extending the convolution targets to mask images, the compact model can adapt to deal with more comprehensive layout information [11]. By refining the convolution kernel design, such as incorporating a Gaussian diffusion function in the z-direction, the compact model can be used to accurately simulate the 3D resist profiles [12]. These enhancements, while effective in specific scenarios, primarily focus on increasing the complexity of the compact models by adding new terms or refining existing components within the framework of linear combinations. Moreover, the growth in model complexity can lead to overfitting or ill-conditioned issues. Current calibration methods, such as genetic algorithms [13], mini-batch gradient descent (MGD) [14], and Landweber iteration method [15] often lack adequate constraints and often require engineers to have extensive experience or resort to trial-and-error over long periods.

In this paper, we go beyond the tradition of resist modeling by a linear combination of model terms, and propose an efficient resist model based on a Wiener–Padé framework, which leverages the advantages of both polynomial and rational approximations. This method not only maintains the computational efficiency but also offers the improved adaptability to complex nonlinear resist responses, providing a versatile solution for a wide range of lithographic applications. In addition, we develop a two-stage calibration method for the Wiener–Padé model based on quadratic convex optimization algorithm. This method integrates the advantages of global optimality, computational efficiency, and constrained solutions, making it both efficient and robust for model tuning. In Section 2, we present the construction of this resist model, and a two-stage calibration approach based on the quadratic convex optimization algorithm. In Section 3, we conduct a comprehensive evaluation of the modeling accuracy and calibration efficiency. The conclusions are drawn in Section 4.

2. Theory and methods

2.1. Wiener–Padé resist modeling

Photoresists exhibit complex nonlinear effects in response to optical intensity. Polynomial approximation is one of the most commonly used methods to model such nonlinear behaviors due to its simplicity and ease of implementation. However, it often struggles with capturing the complex higher-order responses and quickly becomes unwieldy as the degree of the polynomial increases. Padé approximation, a classical method that represents functions as the ratio of polynomials

$$R(x) = \frac{\sum_{i=0}^{m} a_i x^i}{1 + \sum_{j=1}^{n} b_j x^j}, \text{ with } 1 + \sum_{j=1}^{n} b_j x^j > 0, \tag{1}$$

provides an alternative with significant advantages, since it may converge faster and provide better approximations [16–18]. To illustrate, we consider the Padé and polynomial approximation

$$R(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n, \tag{2}$$

of the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-kx}}, \tag{3}$$

one of the simplest resist models [1], for various steepness parameters k and polynomial degrees. Fig. 1 compares the polynomial and Padé approximations with degrees from 2 to 5 (columns) and k values of 0.5, 1, 2, and 10 (rows). It shows that as the polynomial degree increases, the fitting accuracy of both models generally improves, with Padé consistently showing superior performance. Notably, in each k value (row), the Padé approximation with a lower degree often outperforms

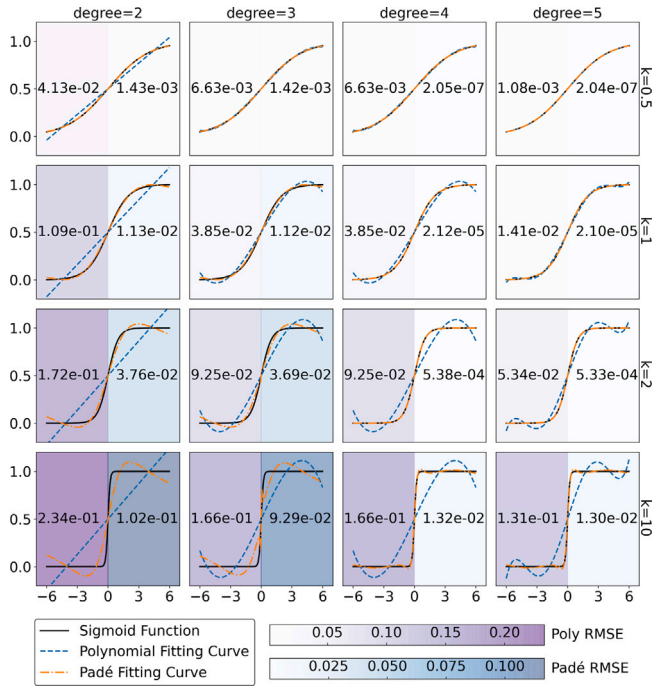

Fig. 1. Comparison of polynomial and Padé approximations using sigmoid functions. (RMSE heatmaps for the polynomial and Padé approximations are displayed on the left and right halves of each subplot, respectively).

the polynomial with a higher degree. For instance, at $k = 2$, the Root Mean Square Error (RMSE) for the 5th-degree polynomial is 5.34e−02, whereas the RMSE for the 2nd-degree Padé is 3.76e−02, indicating that the Padé approximation can achieve better accuracy with a lower complexity. Although the simplified sigmoid response used for illustration does not encompass the full range of physicochemical effects in resist behavior (such as resist shrinkage and developer diffusion), the Padé approximation's superior mathematical flexibility in approximating complex shapes remains evident.

The Wiener model is also a mathematical model used to describe nonlinear systems, combining the polynomial functions and the convolution operations to capture the relationship between inputs and outputs [19–22]. It is particularly suitable for modeling photoresists due to their inherent nonlinear response characteristics. Its flexibility and adjustable parameters make it well-adapted to various materials and conditions in lithography processes [23]. Formally, the Wiener model expresses the output as:

$$R(I) \approx \sum_{n=1}^{N} \left[\sum_{\{k_i\}=1}^{K} w_{k_1 k_2 \ldots k_n} \left(H_{k_1} \otimes I\right)\left(H_{k_2} \otimes I\right) \cdots \left(H_{k_n} \otimes I\right) \right], \tag{4}$$

where H_k are the Wiener kernels, I is the input aerial image, and $w_{k_1 k_2 \ldots k_n}$ are the Wiener coefficients. If we define $(H_{k_n} \otimes I)$ as the linear Wiener product (WP), $(H_{k_n} \otimes I)(H_{k_{n+1}} \otimes I)$ as the quadratic WP, $(H_{k_n} \otimes I)(H_{k_{n+1}} \otimes I)(H_{k_{n+2}} \otimes I)$ as the cubic WP, and so on, then by linearly combining all these terms, we obtain:

$$R(I) = \sum_i w_i \mathrm{WP}_i(I), \tag{5}$$

It can be seen that the Wiener model combines the WPs through a linear combination like the polynomial. By combining Padé approximations and the Wiener model framework, we propose a Wiener–Padé model

as [24]:

$$R(I) = \frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I)}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I)}, \text{ with } 1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I) > 0, \tag{6}$$

where Z is the number of numerator Wiener terms, w_i and w_j are Wiener coefficients, $\mathrm{WP}_i(I)$ and $\mathrm{WP}_j(I)$ are Wiener products. This way, the Wiener–Padé model offers several advantages in resist modeling:

- Efficient approximation capability: The Wiener–Padé model combines the strengths of the Wiener model in describing nonlinear relationships and the efficiency of Padé approximation in approximating complex functions with fewer lower-order terms. Compared to models that rely solely on linear combinations, the Wiener–Padé model can achieve effective approximation of complex nonlinear resist responses with lower model orders.
- Better representation of nonlinear behavior: Rational functions in the Wiener–Padé model are generally non-divergent, making them particularly well-suited for modeling complex responses within resist. In contrast, conventional compact models based on linear combinations may require higher-order terms to describe these behaviors, leading to increased computational cost.
- Reduced computational complexity in optimization: The Wiener–Padé model can achieve homogeneous terms by utilizing specific combinations. Such characteristic avoids the need for taking roots of high-power model terms, which significantly reduces the computational complexity involved in gradient-based optimization.
- Reduced overfitting risk: The Wiener–Padé model uses a rational function to approximate complex behaviors, effectively avoiding overfitting, especially in cases with limited data, such as the currently widely used calibration scenarios using CD gauges or contour points.
- High flexibility and versatility: By flexibly adding and combining terms in the numerator and denominator, it can adapt to a wide range of resist modeling scenarios. Traditional compact models can be regarded as a simplified linear form of the general Wiener–Padé model.

2.2. Wiener–Padé model calibration

For calibration, a resist model typically requires a large amount of actual CD measurements where genetic algorithm is commonly employed to ensure the global optimal solution. In order to avoid excessive computational time., gradient-based approaches, such as the MGD and the projected Landweber iteration algorithm, have been proven effective. Nevertheless, the performance and convergence speed of such methods are highly dependent on the accuracy of the initial point. In particular, poor initial values can lead to slow convergence. Additionally, due to the ill-conditioned problem and overfitting issues induced by the increased complexity of resist models, incorporating appropriate constraints into the calibration process becomes crucial as well.

To address these issues, we propose a two-stage approach based on the quadratic convex optimization algorithm [25–28]:

1. The qualitative calibration stage
   The goal of this stage is to find a global approximate solution such that the simulated values of most CD gauges are close to the measured values. By leveraging the convex structure of the problem, we solve a relaxed version of the optimization to generate an initial guess. This step avoids local optima and accelerates convergence in subsequent iterations (see Appendix A.1 for mathematical formulation).
2. The quantitative calibration stage
   Starting from the initial values obtained in the qualitative stage, we refine the solution through iterative optimization. We reconstruct the objective function in each iteration to minimize the discrepancy between the Wiener–Padé model simulations and actual measurement data by searching for the optimal perturbation. This allows for progressive refinement of the solution, offering advantages such as robustness and the ability to converge to a highly accurate solution even with complex model landscapes (algorithm details in Appendix A.2).

In both processes, we apply appropriate constraints to prevent overfitting. With the proposed strategy, we can improve model accuracy while reducing calibration time.

3. Experimental results

We have conducted a series of experiments to demonstrate the advantages of the proposed Wiener–Padé modeling and calibration method. We first evaluated the modeling accuracy of the Wiener–Padé model, then compared the accuracy and efficiency of different calibration methods. Finally, we compared the calibration results for two practical lithographic models using the complete Wiener–Padé modeling and calibration approach versus the Wiener model.

3.1. Performance of a Wiener–Padé model

To validate the theoretical feasibility of a Wiener–Padé model, we conducted a series of experiments to compare the Wiener and proposed Wiener–Padé models' fitting performances. We used a simplified Wiener model of the following form:

$$R(I) = a_0 + a_1 (H \otimes I) + a_2 (H \otimes I)^2 + \cdots + a_n (H \otimes I)^n, \tag{7}$$

and a Wiener–Padé model as follows:

$$R(I) = \frac{\sum_{i=0}^{n} a_i (H \otimes I)^i}{1 + \sum_{j=0}^{n} b_j (H \otimes I)^j}, \text{ with } 1 + \sum_{j=0}^{n} b_j (H \otimes I)^j > 0, \tag{8}$$

where, H is a Gaussian kernel with $\sigma = 30$ nm was used; I is generated under annular illumination ($\sigma_{in} = 0.5, \sigma_{out} = 0.75$) in an immersion lithography system with $\lambda = 193$ nm and NA = 1.2.

There are two typical mask patterns: an array of equal lines and spaces and an isolated line or space [29].

Ex.1. Dense Array of Lines and Spaces

For periodic dense lines and spaces, the spatial frequency response of the diffraction pattern appears as discrete diffraction orders [29]:

$$T_{dense}(f_x) = \frac{1}{p} \sum_{j=-\infty}^{\infty} \frac{\sin(\pi w f_x)}{\pi f_x} \delta\left(f_x - \frac{j}{p}\right), \tag{9}$$

where, δ is the Dirac delta function, which is the mathematical representation of a point of light. The final aerial image is formed by the incoherent superposition of coherent images corresponding to all discrete source points, where each coherent image is derived by taking the inverse Fourier transform of the shifted spatial frequency response $T(f_x)$ and then squaring its magnitude. In addition, we defined the ideal resist image using the sigmoid function [1]:

$$R_{dense}(x) = \sigma(\cos(2\pi x/p)) = \frac{1}{1 + e^{-k\cos(2\pi x/p)}}. \tag{10}$$

In this experiment, the aerial image was normalized, and the following parameters were applied:

x: Horizontal position, varied from −320 nm to 320 nm with 640 sampling points;

w: Linewidth, set to 80 nm;

p: Line space pitch, set to 160 nm;

k: Slope factor of sigmoid function, set to 10.

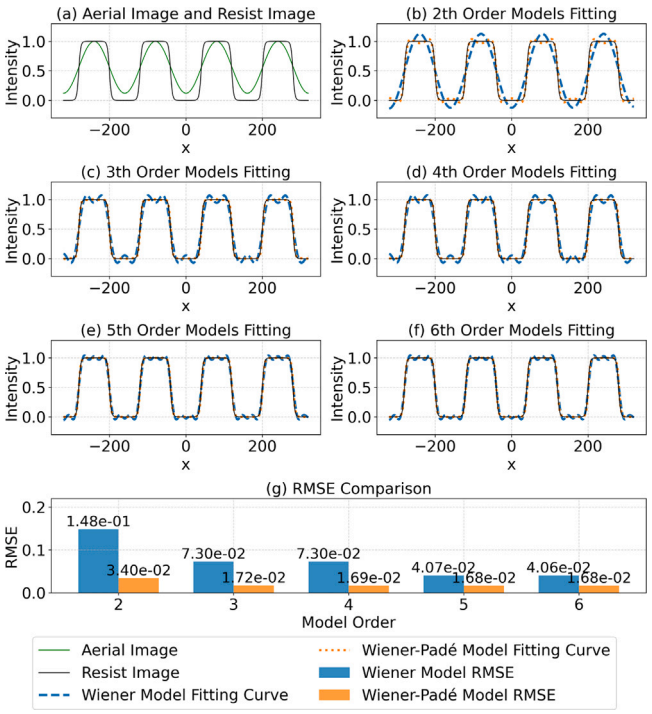

Fig. 2. Model fitting comparison for a dense array of lines and spaces.

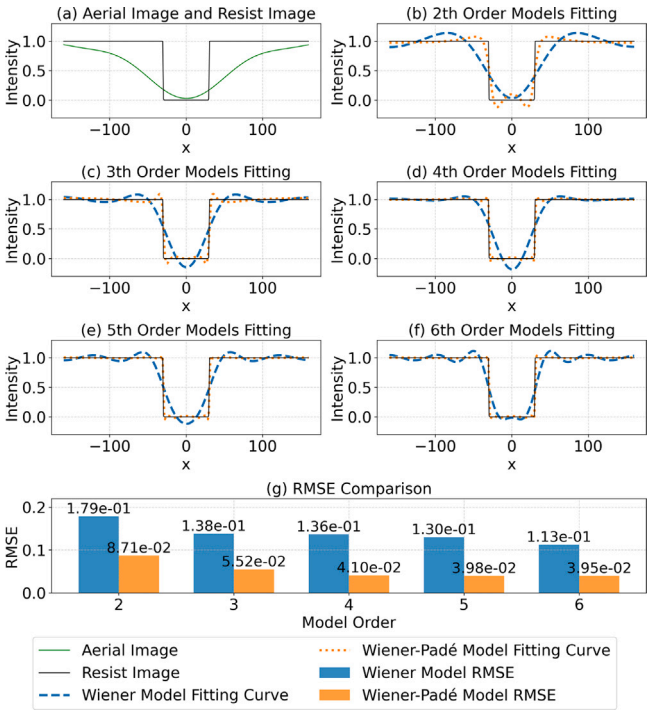

Fig. 3. Model fitting curve comparison for an isolated line or space.

We employed the Levenberg–Marquardt algorithm, a commonly used nonlinear least squares method, to fit the models. The sets of RMSE between the target and the simulation were computed over a defined interval of x values for comparison and evaluation. Fig. 2 demonstrates the fitting performance of Wiener and Wiener–Padé model for the resist image of a dense array of lines and spaces across different model orders, along with a comparison of RMSE values. Fig. 2(a) shows the input aerial image and corresponding output resist image. Figs. 2(b) to 2(f) illustrate the fitting results for model orders from 2 to 6, where the blue dashed lines represent the Wiener model and the orange dash-dotted lines represent the Wiener–Padé model. Fig. 2(g) provides a comparison of RMSE values for each model order, showing that the Wiener–Padé model consistently achieves lower RMSE than the Wiener model. Notably, the RMSE of the 2nd order Wiener–Padé model (3.40e−02 nm) is lower than that of the 6th order Wiener model (4.06e−02 nm).

Ex.2. Isolated Line or Space

For an isolated space, its diffraction pattern's spatial frequency response manifests as a sinc function [29]:

$$T_{iso}(f_x) = \frac{\sin \pi w f_x}{\pi f_x}. \tag{11}$$

We then defined the ideal resist profile using the sigmoid function [1]:

$$R_{iso}(x) = 1 - \left(\frac{1}{1 + e^{-k(x-c_1)}} + \frac{1}{1 + e^{k(x-c_2)}} \right). \tag{12}$$

In this experiment, the aerial image was also normalized, and the following parameters were applied:

x: Horizontal position, varied from −160 nm to 160 nm with 320 sampling points;

w: Width of the isolated line or space, set to 80 nm;

k: Slope factor of sigmoid function, set to 10;

c_1: Left transition point, set to −40 nm;

c_2: Right transition point, set to 40 nm.

Fig. 3, similar to the previous Fig. 2, illustrates the fitting performance and comparison of Wiener and Wiener–Padé models for resist image intensity of an isolated line or space. The nonlinearity of this pattern is slightly higher compared to the dense array of lines and spaces, which is reflected in the overall results. However, the results still reveal a consistent trend where the Wiener–Padé model significantly outperforms the Wiener model across all orders. Notably, the 2nd order Wiener–Padé model achieves an RMSE of 8.71e−02 nm, also lower than the 6th order Wiener model's 1.13e−01 nm.

These two comparative analyses highlight the significant advantages of the Wiener–Padé model in enhancing nonlinear resist modeling. By integrating Padé approximation into the Wiener model, higher accuracy with lower computational complexity can be achieved.

3.2. Performance of two-stage quadratic convex optimization-based calibration

To assess the accuracy and efficiency of the two-stage quadratic convex optimization approach, we conducted a series of experiments using a real DUV immersion lithography case. The model parameters included an illumination wavelength of 193 nm, an NA of 1.2, and a 6% attenuated phase shift mask (PSM). The optical setup utilized an annular illumination source with $\sigma = 0.5/0.75$ and XY-polarization. The wafer stack consisted of a tri-layer structure, from top to bottom: an 85-nm-thick JSR A2055 photoresist, a silicon-containing anti-reflective coating (SiARC), and a spin-on carbon (SOC) layer. Focus settings covered a ±60 nm defocus range with 20-nm steps. Processing involved a post-exposure bake (PEB) at 105 °C for 60 s. The test patterns, with feature sizes ranging from 40 nm to 500 nm, comprised over 3000 different test patterns, from which we randomly selected 100 CD gauges. Using a Wiener–Padé model with two linear terms and six quadratic terms, we calibrated the model parameters using three different methods: the classical genetic algorithm, the projected Landweber method, and the two-stage quadratic convex optimization approach. The experiments were implemented in C++ and executed using OpenMPI. We recorded the calibration runtimes and the RMSEs in one slave rank on hardware consisting of an Intel Core i7-10700 CPU, which features 8 cores and 16 threads, a base frequency of 2.90 GHz, and a maximum

Table 1
Comparison of calibration accuracy and efficiency across different methods.

Model	RMSE (nm)	Time (h)
Classical genetic algorithm	0.989555	25.18
Projected Landweber method	1.050881	0.61
Two-stage approach	0.990187	0.02

turbo boost of 4.80 GHz. The average memory consumption during optimization was measured as 453 MB for the Landweber method, 466 MB for the two-stage approach, and 516 MB for the genetic algorithm, with all methods showing negligible memory fluctuations. The data presented in Table 1 shows that the two-stage quadratic convex optimization approach achieved an RMS of 0.990187 nm, which was on par with the genetic algorithm (0.989555 nm) but significantly better than the Landweber method (1.050881 nm). In terms of efficiency, the two-stage approach demonstrated a substantial improvement, requiring only 0.02 h for calibration, compared to 25.18 h for the genetic algorithm and 0.61 h for the Landweber method. These results indicate that the two-stage approach not only maintains the accuracy of traditional methods like the genetic algorithm but also significantly reduces the computational time, making it a more efficient and practical option for resist calibration.

We also observed that, in the absence of constraints, both the genetic algorithm and the Landweber method occasionally encountered several issues, as illustrated by the green cutlines in Fig. 4. First of all, Fig. 4(a) shows very low intensity contrast, resulting in an extremely narrow process window. Additionally, while the intensity profile intersects the threshold, the polarity of the intensity inside the contour in Fig. 4(b) is reversed. Furthermore, Fig. 4(c) demonstrates multiple intersections between the cutline and the threshold, resulting in the formation of holes. In Fig. 4(d), although the polarity of the intensity is correct, the intensity outside the contour approaches the threshold at distant locations, which can potentially result in the formation of multiple potential contours. These issues are difficult to detect solely based on the simulated CD values but can significantly impact the model's applicability in large-area simulations and process window analysis. This highlights the importance of incorporating constraints on the intensity distribution in resist model calibration.

To further evaluate the effect of constraints on the resist intensity distribution, we conducted an experiment on overfitting issues. Using these gauges, we calibrated a Wiener model with 38 model terms to investigate the impact of different constraint conditions on the calibration results. Specifically, we compared two scenarios: heavily constrained (red cutlines in Fig. 4) and lightly constrained conditions (yellow cutlines in Fig. 4), both employing a 1 nm sampling interval (indicated by dark red dots) and a 0.2 threshold value. Under heavily constrained conditions, constraints were applied to all sampling points across the entire CD gauge cutline between the two endpoints. In contrast, in the lightly constrained case, constraints were selectively applied only to a specifically chosen subset of sampling points, comprising five key locations: the two ends of the cutline, the center point, and two positions near the measurement locations. It is important to note that the manner of constraint application remained consistent across both scenarios. Taking Fig. 4(a) as an example, for sampling points where the simulated intensity was expected to be near the threshold, constraints were applied according to Eq. (23). For points expected to be below the threshold, constraints were implemented as described in Eq. (24), while for points expected to be above the threshold, constraints were formulated based on Eq. (25). The results, as depicted in Fig. 4, indicate that both types of constraints effectively addressed the overfitting issues observed under unconstrained conditions. Notably, heavily constrained conditions led to higher intensity contrast and imaging quality, while lightly constrained conditions kept all intensity distributions within the constraint range, making the intensity values more controllable. Therefore, appropriate constraints on the resist intensity distribution are crucial in the calibration process.

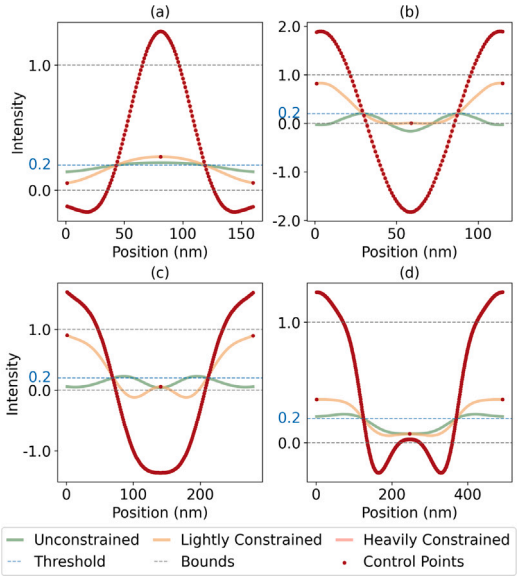

Fig. 4. Intensity distribution of resist response for different constraint conditions along the CD gauges: (a) 80 nm spaces with 170 nm pitch, (b) 48 nm lines with 110 nm pitch, (c) a 45 nm line between two 74 nm space ends, (d) a 160 nm space between dense 70 nm line ends with 126 nm pitch. Each subplot shows cutlines of the simulated resist response after calibration, illustrating issues encountered in unconstrained optimization (green lines) and the effects of heavily constrained (red lines) and lightly constrained (yellow lines) conditions. Dark red dots indicate the control points used in the constrained cases.

3.3. A holistic lithography test case

To test the effectiveness of the Wiener–Padé modeling and calibration methods in practical applications, we conducted comparative experiments using a positive tone development (PTD) and a negative tone development (NTD) lithographic process. In the PTD case, the lithographic model parameters were the same as introduced in Section 3.2, but with 3271 gauges used for calibration and 164 gauges used for verification. For the NTD model, the parameters were $\lambda = 193$ nm, NA = 1.3, and a 6% attenuated PSM mask with a −0.5 nm pattern bias. In this case, 1500 gauges were used for calibration and 93 gauges for verification.

Table 2 shows the comparative performance metrics for the Wiener and Wiener–Padé models in the PTD and NTD cases, where model calibration is considered complete when 95% of the simulation errors for the gauges are within the reference range. For the PTD case, the reference range is ±2.5 nm for 1D patterns and ±6 nm for 2D patterns. Similarly, for the NTD case, the reference range is ±2 nm for 1D patterns and ±4 nm for 2D patterns. We utilized a unified set of terms, composed of aerial image features and Laguerre–Gauss kernels, for both PTD and NTD calibrations. It is evident that the Wiener–Padé model achieves calibration using only approximately two-thirds of the term count required by the Wiener model, while also demonstrating superior RMSE and error range metrics. Specifically, for the PTD case, the Wiener–Padé model achieves an RMSE of 0.8784 nm compared to 0.9523 nm for the Wiener model, with an error range of 15.4820 nm versus 17.5470 nm for the calibration gauges. For the verification gauges, the RMSE is 0.8873 nm for the Wiener–Padé model, outperforming the Wiener model's 1.0505 nm, with a reduced error range of 4.3942 nm compared to 6.9559 nm. In the NTD case, the Wiener–Padé model achieves an RMSE of 1.1754 nm (vs. 1.2834 nm) and an error range of 13.3772 nm (vs. 16.6335 nm) for the calibration gauges. For the verification gauges, it maintains its advantage with an RMSE of 5.1918 nm compared to 5.3909 nm and an error range of 32.1864 nm versus 33.7054 nm.

Table 2
Comparative performance metrics for the Wiener and the Wiener–Padé model.

Case	Model	Term	Calibration			Verification	
			RMSE (nm)	Error range (nm)	Within range (%)	RMSE (nm)	Error range (nm)
PTD	Wiener	13	0.9523	17.5470	97.03	1.0505	6.9559
	Wiener–Padé	8	0.8784	15.4820	98.07	0.8873	4.3942
NTD	Wiener	19	1.2834	16.6335	95.00	5.3909	33.7054
	Wiener–Padé	12	1.1754	13.3772	95.53	5.1918	32.1864

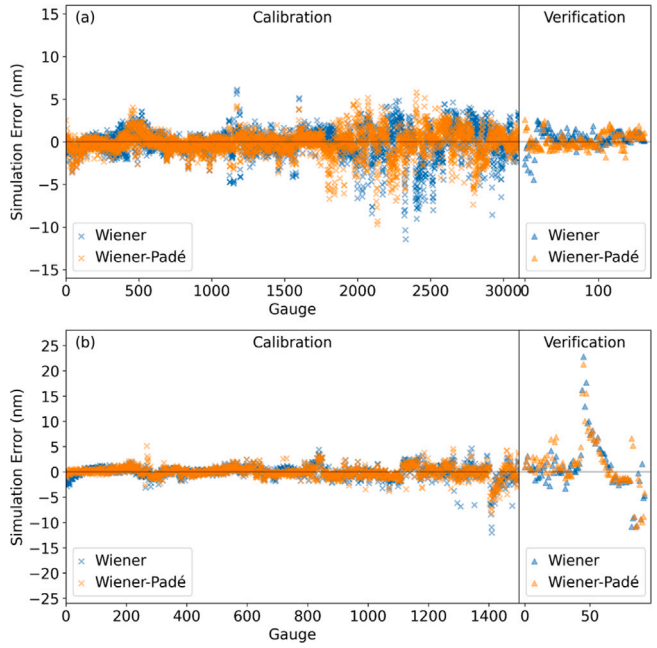

Fig. 5. Distribution of simulation errors for the Wiener and the Wiener–Padé Model in the (a) PTD and (b) NTD lithographic cases.

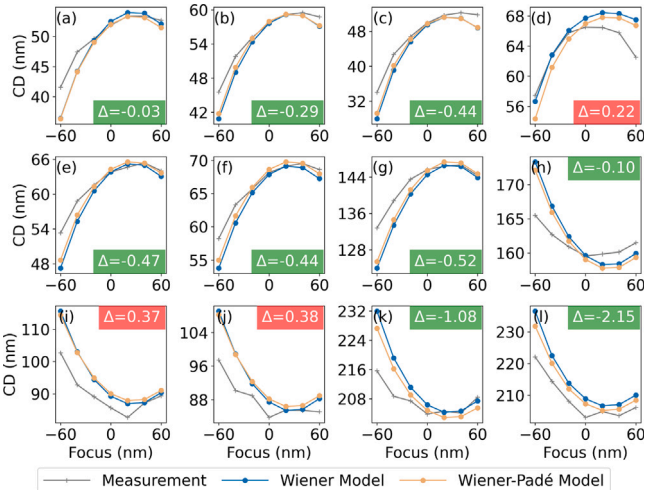

Fig. 6. Bossung plots of the Wiener and the Wiener–Padé Model in the PTD lithographic case under different defocus conditions for patterns: (a) 64 nm lines with 260 nm pitch, (b) 64 nm lines with 1500 nm pitch, (c) 58 nm lines with 116 nm pitch, (d) 80 nm spaces with 160 nm pitch, (e) 74 nm lines with 340 nm pitch, (f) 74 nm lines with 1000 nm pitch, (g) 140 nm lines with 280 nm pitch, (h) 160 nm spaces with 320 nm pitch, (i) a 60 nm space between line ends with 80 nm width, (j) a 70 nm space between line ends with 100 nm width, (k) 150 spaces nm between line ends with 60 nm width and 126 nm pitch, and (l) 150 nm spaces between line ends with 70 nm width and 200 nm pitch.

Fig. 5 further illustrates these trends. For both the PTD and the NTD cases, the Wiener–Padé model demonstrates more concentrated errors with better-controlled error ranges in the calibration part. For the PTD verification patterns, the Wiener model shows a subset of errors that deviate significantly, while the Wiener–Padé model maintains errors tightly distributed around zero. In the NTD verification tests, both the Wiener and the Wiener–Padé models struggled to squeeze the model errors down to a desired $\sim$ 10 nm level, indicating either a resist response beyond the employed Wiener and Wiener–Padé model forms or insufficient pattern coverage in the set of selected training data. One possibility is to incorporate a resist shrinkage model beyond the Wiener or Wiener–Padé framework. Another approach is to enrich the training data to include some of the typical failing patterns. Both directions are being considered and investigated in our ongoing research.

To assess the generality and applicability of the Wiener–Padé model, we conducted a process window analysis. Figs. 6 and 7 illustrate the Bossung curves of the Wiener and the Wiener–Padé model in the PTD and NTD cases under different defocus conditions for different Focus Exposure Matrix (FEM) patterns. In addition, we calculated the mean absolute errors (MAE) between simulated and measured values for each model across defocus. The difference in MAE, denoted as Δ, is visually represented in Figs. 6 and 7. Values with red backgrounds indicate cases where the Wiener–Padé model exhibits larger MAE than the Wiener model, while values with green backgrounds signify superior Wiener–Padé performance. For the PTD case, the patterns are: (a) 64 nm lines with 260 nm pitch, (b) 64 nm lines with 1500 nm pitch, (c) 58 nm lines with 116 nm pitch, (d) 80 nm spaces with 160 nm pitch, (e) 74 nm lines with 340 nm pitch, (f) 74 nm lines with 1000 nm pitch, (g) 140 nm lines with 280 nm pitch, (h) 160 nm spaces with 320 nm pitch, (i) a 60 nm space between line ends with 80 nm width, (j) a 70 nm space between line ends with 100 nm width, (k) 150 spaces nm between line ends with 60 nm width and 126 nm pitch, and (l) 150 nm spaces between line ends with 70 nm width and 200 nm pitch. The defocus range spans from −60 nm to +60 nm with a 20 nm interval. As shown in Fig. 6, the Wiener–Padé model achieves smaller MAE in 9 out of 12 patterns, providing more accurate predictions. For the NTD case, the patterns are: (a) 74 nm lines with 700 nm pitch, (b) 90 nm lines with 600 nm pitch, (c) 106 nm lines with 100 nm pitch, (d) an isolated 73 nm line, (e) two 150 nm lines, (f) three 150 nm lines, (g) 90 nm lines with 400 nm pitch, (h) 150 nm lines with 430 nm pitch, (i) 64 nm lines with 600 nm pitch and scattering bars, (j) 64 nm lines with 1500 nm pitch and scattering bars, (k) 120 spaces nm between line ends with 70 nm width and 140 nm pitch, and (l) 60 nm spaces between line ends with 74 nm width and 148 nm pitch. The defocus values range from −100 nm to 100 nm. Fig. 7 also demonstrates the superior performance of the Wiener–Padé model, with 8 out of 12 patterns exhibiting smaller MAE compared to the Wiener model. Furthermore, for cases such as (i) and (j), which include scattering bars, and (k) and (l), which are 2D patterns, the Wiener–Padé model shows a significantly closer trend than the Wiener model. These indicate that the Wiener–Padé model has better predictive capability for the process window.

To evaluate the model complexity and computational efficiency, we recorded the combination of terms and the calibration time for each model, as shown in Table 3. In both PTD and NTD cases, the Wiener–Padé model uses significantly fewer terms than the Wiener model (6 vs. 12 for PTD, 10 vs. 18 for NTD), especially quadratic terms, which contribute more to the model's complexity. In addition, the Wiener–Padé model completes calibration in 0.11 h, significantly faster than

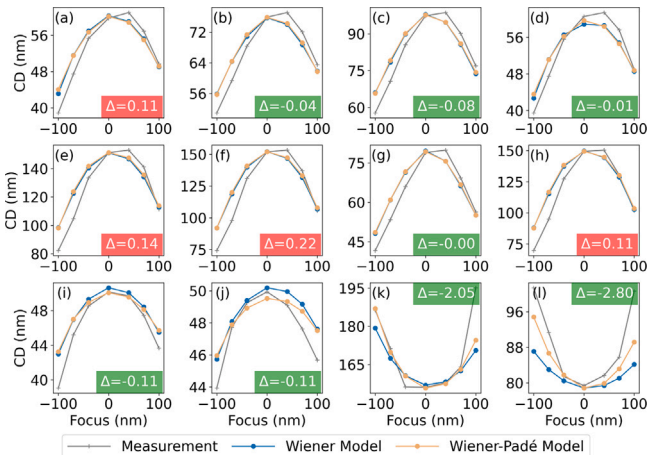

Fig. 7. Bossung plots of the Wiener and the Wiener–Padé Model in the NTD lithographic case under different defocus conditions for patterns: (a) 74 nm lines with 700 nm pitch, (b) 90 nm lines with 600 nm pitch, (c) 106 nm lines with 100 nm pitch, (d) an isolated 73 nm line, (e) two 150 nm lines, (f) three 150 nm lines, (g) 90 nm lines with 400 nm pitch, (h) 150 nm lines with 430 nm pitch, (i) 64 nm lines with 600 nm pitch and scattering bars, (j) 64 nm lines with 1500 nm pitch and scattering bars, (k) 120 spaces nm between line ends with 70 nm width and 140 nm pitch, and (l) 60 nm spaces between line ends with 74 nm width and 148 nm pitch.

Table 3
Comparative calibration time for the Wiener and the Wiener–Padé model.

Case	Model	Terms (Linear+Quadratic)	Calibration time (h)
PTD	Wiener	13 (1+12)	1.81
	Wiener–Padé	8 (2+6)	0.11
NTD	Wiener	19 (1+18)	0.87
	Wiener–Padé	12 (2+10)	0.03

the Wiener model's 1.81 h. For the NTD case, due to a reduction in the number of calibration gauges from over 3000 to 1500, both models require less calibration time, but the Wiener–Padé model shows a much greater reduction, completing calibration in just 0.03 h compared to the Wiener model's 0.87 h.

4. Conclusion

In this paper, we presented a method for fast and accurate resist modeling using a Wiener–Padé formulation and a means of two-stage calibration using quadratic convex optimizations with effective constraints, aimed at improving the adaptability of compact models in handling complex nonlinear resist responses as well as addressing the challenges of runtime efficiency and overfitting in calibration.

Compared with the Wiener model, the present Wiener–Padé model showed superior fitting accuracy with lower model order. Particularly, the 2nd-order Wiener–Padé model exhibited superior performance than the 6th-order Wiener model. The two-stage quadratic convex optimization-based calibration approach exhibited comparable accuracy as genetic algorithm and over 30 times efficiency upon the Landweber algorithm. Furthermore, we showcased the complete Wiener–Padé modeling and calibration method using the compact Wiener model as baseline in a practical PTD and a NTD case. The Wiener–Padé model outperformed the Wiener model across various performance metrics in both two cases, including RMSE, error range, and process window prediction, for calibration and verification patterns. Additionally, the Wiener–Padé model used only around 2/3 terms required by the Wiener model in both two cases, and the calibration time represented a reduction of at least over 16 times. These results demonstrate that the Wiener–Padé model, combined with the proposed two-stage calibration approach, provides a promising solution for advanced resist modeling and calibration in computational lithography, with higher accuracy and efficiency as well as reduced complexity.

Funding

National Natural Science Foundation of China (52205592, 52130504), Key Research and Development Plan of Hubei Province (2022BAA013), Major Program (JD) of Hubei Province (2023BAA008-2), and Innovation Project of Optics Valley Laboratory (OVL2023PY003).

CRediT authorship contribution statement

Chunxiao Mu: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhiyang Song:** Writing – review & editing, Validation, Investigation, Data curation. **Lei Cheng:** Writing – review & editing, Validation, Investigation, Data curation. **Shaopeng Guo:** Writing – review & editing, Validation. **Ke Li:** Writing – review & editing, Validation. **Song Zhang:** Writing – review & editing, Validation, Funding acquisition. **Hao Jiang:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **David H. Wei:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Yanlong Sun:** Writing – review & editing, Methodology. **Jinlong Zhu:** Writing – review & editing. **Shiyuan Liu:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hao Jiang reports financial support was provided by Department of Science and Technology of Hubei Province. Shiyuan Liu reports financial support was provided by National Natural Science Foundation of China. Song Zhang reports financial support was provided by National Natural Science Foundation of China. David H. Wei has patent issued to Yuwei Optica Co.,Ltd, Wuhan. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their thanks to the technical support from the Experiment Centre for Advanced Manufacturing and Technology in School of Mechanical Science and Engineering of HUST.

Appendix. Mathematical details of the two calibration stages

A.1. The qualitative calibration stage

As shown in Fig. 8, the primary goal in this stage is to ensure that the simulation values at measured positions p closely align with the model threshold TH. Beyond this, the simulation values at points inside and outside the contour should deviate from the threshold as much as possible. Depending on the polarity of the test patterns, our point selection varies. As illustrated in Fig. 8(a), for a mask with bright-field, points that deviate downward from the threshold $p^{(-)}$ are inside the contour, while those that deviate upward $p^{(+)}$ are outside. For the pattern in Fig. 8(b) with opposite polarity, the point selection strategy is reversed. Lastly, all simulation values on the CD gauges should remain within a constraint range.

Specifically, we extract the simulated values based on the coordinate positions determined from the actual measured contours or CD

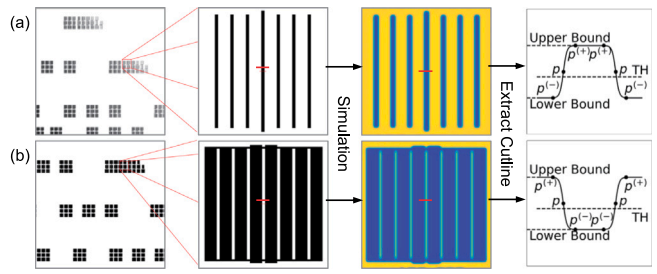

Fig. 8. Constraint settings for test patterns of different polarities. Both (a) and (b) illustrate the same process: starting from the mask layout where the respective test pattern (line for (a), space for (b)) is selected, followed by resist image simulation, and finally cutline extraction with corresponding constraint settings applied to the resist image.

data. The optimization objective $f_{(0)}$ and constraint for the simulated values on the contours is defined as:

$$\min f_{(0)} = \sum_{m=1}^{M} \left(R(I_{p_m}) - \mathrm{TH} \right)^2, \text{s.t. } L \le R(I_{p_m}) \le U \tag{13}$$

where, $R(I_{p_m})$ represents the Wiener–Padé output at the point p_m, TH is the model threshold, L is the lower bound, U is the upper bound. Additionally, we sample the simulated values from regions both interior and exterior to the contours. The optimization objectives for threshold-boundary constrained areas, denoted as $f_{(-)}$ and $f_{(+)}$, are defined as follows: For regions where values should remain below the threshold TH (e.g., exterior points in Fig. 8(a) or interior points in Fig. 8(b)), we define $f_{(-)}$ to penalize values approaching or exceeding TH. Conversely, for regions requiring values to stay above TH (e.g., interior points in Fig. 8(a) or exterior points in Fig. 8(b)), we formulate $f_{(+)}$ to penalize values approaching or falling below TH. Specifically, these terms are expressed as follows:

$$\min f_{(-)} = \sum_{m^{(-)}=1}^{M^{(-)}} \left(R(I_{p_m^{(-)}}) - \mathrm{TH} \right), \text{s.t. } L \le R(I_{p_m^{(-)}}) \le \mathrm{TH} \tag{14}$$

$$\min f_{(+)} = \sum_{m^{(+)}=1}^{M^{(+)}} \left(\mathrm{TH} - R(I_{p_m^{(+)}}) \right), \text{s.t. } \mathrm{TH} \le R(I_{p_m^{(+)}}) \le U \tag{15}$$

where, $R(I_{p_m^{(-)}})$ and $R(I_{p_m^{(+)}})$ represent the Wiener–Padé outputs at the points $p_m^{(-)}$ and $p_m^{(+)}$, respectively.

Given all WP terms, the resist model calibration essentially becomes a quadratic convex optimization problem:

$$\min \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x}, \text{s.t. } \mathbf{l} \le \mathbf{C}\mathbf{x} \le \mathbf{u}. \tag{16}$$

where, $\mathbf{x}$ is the parameter vector to be optimized, which is expressed as $\mathbf{x} = [w_1, w_2, w_3, \ldots, w_n, \mathrm{TH}]^T$.

The quadratic term is therefore expressed as:

$$\begin{aligned}
&\min \varepsilon_g = f_{(0)} \\
&= \sum_{m=1}^{M} \left(\frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m})} - \mathrm{TH} \right)^2 \\
&= \sum_{m=1}^{M} \left(\sum_{n=1}^{N+1} w_n \mathrm{WP}_n(I_{p_m})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right)^2 \\
&= \sum_{u,v=1}^{N+1} \left[\sum_{m=1}^{M} \begin{pmatrix} \mathrm{WP}_u(I_{p_m})(\mathbf{1}_{u\le Z} - \mathrm{TH}(1 - \mathbf{1}_{u\le Z})) \\ \cdot \mathrm{WP}_v(I_{p_m})(\mathbf{1}_{v\le Z} - \mathrm{TH}(1 - \mathbf{1}_{v\le Z})) \end{pmatrix} \right] w_u w_v \\
&= \mathbf{x}^T\mathbf{P}\mathbf{x},
\end{aligned} \tag{17}$$

and

$$\mathbf{1}_{n\le Z} = \begin{cases} 1, & \text{if } n \le Z, \\ 0, & \text{if } n > Z. \end{cases} \tag{18}$$

The matrix $\mathbf{P} \in \mathbb{R}^{(N+1)\times(N+1)}$ is a positive semidefinite matrix:

$$[\mathbf{P}]_{uv} = \sum_{m=1}^{M} \begin{pmatrix} \mathrm{WP}_u(I_{p_m})(\mathbf{1}_{u\le Z} - \mathrm{TH}(1 - \mathbf{1}_{u\le Z})) \\ \cdot \mathrm{WP}_v(I_{p_m})(\mathbf{1}_{v\le Z} - \mathrm{TH}(1 - \mathbf{1}_{v\le Z})) \end{pmatrix}. \tag{19}$$

We introduce the matrix $\mathbf{W} \in \mathbb{R}^{M\times(N+1)}$, and define it as:

$$[\mathbf{W}]_{m,n} = \mathrm{WP}_n(I_{p_m}) \left(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z}) \right). \tag{20}$$

For each p, it holds that $\mathrm{WP}_{n+1}(I_{p_m}) = -1$. Thus, $\mathbf{P} = \mathbf{W}^T\mathbf{W}$ and $\varepsilon_q = \|\mathbf{W}\mathbf{x}\|^2$.

The linear term $\mathbf{q}^T\mathbf{x}$ is then formulated by combining the optimization terms $f_{(-)}$ and $f_{(+)}$:

$$\begin{aligned}
&\min \varepsilon_l = f_{(-)} + f_{(+)} \\
&= \sum_{m^{(-)}=1}^{M^{(-)}} \left(\frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m^{(-)}})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m^{(-)}})} - \mathrm{TH} \right) \\
&+ \sum_{m^{(+)}=1}^{M^{(+)}} \left(\mathrm{TH} - \frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m^{(+)}})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m^{(+)}})} \right) \\
&= \sum_{m^{(-)}=1}^{M^{(-)}} \left(\sum_{n=1}^{N+1} w_n \mathrm{WP}_n(I_{p_m^{(-)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right) \\
&+ \sum_{m^{(+)}=1}^{M^{(+)}} \left(- \sum_{n=1}^{N+1} w_n \mathrm{WP}_n(I_{p_n^{(+)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right) \\
&= \sum_{n=1}^{N+1} \left[\sum_{m^{(-)}=1}^{M^{(-)}} \left(\mathrm{WP}_n(I_{p_m^{(-)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right) \right. \\
&\left. - \sum_{m^{(+)}=1}^{M^{(+)}} \left(\mathrm{WP}_n(I_{p_m^{(+)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right) \right] w_n \\
&= \mathbf{q}^T\mathbf{x},
\end{aligned} \tag{21}$$

where, $\mathbf{q} \in \mathbb{R}^{(N+1)}$ is defined as a linear vector:

$$\begin{aligned}
[\mathbf{q}]_n = &\sum_{m^{(-)}=1}^{M^{(-)}} \left(\mathrm{WP}_n(I_{p_m^{(-)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right) \\
&- \sum_{m^{(+)}=1}^{M^{(+)}} \left(\mathrm{WP}_n(I_{p_m^{(+)}})(\mathbf{1}_{n\le Z} - \mathrm{TH}(1 - \mathbf{1}_{n\le Z})) \right).
\end{aligned} \tag{22}$$

For each p, $\mathrm{WP}_{N+1}(I_{p^{(-)}}) = 0$, $\mathrm{WP}_{N+1}(I_{p^{(+)}}) = 0$. The number of points on the internal and external contours can differ, i.e., $M^{(+)} \neq M^{(-)}$.

In the constraint terms, we confine the simulated values $[\mathbf{C}\mathbf{x}]$ at and around the target contour within the convex set $[\mathbf{l}, \mathbf{u}]$, where the lower bound $\mathbf{l} \in \mathbb{R}^{M^*}$, the upper bound $\mathbf{u} \in \mathbb{R}^{M^*}$, the constraint matrix $\mathbf{C} \in \mathbb{R}^{M^*\times(N+1)}$, with M^* is the number of specific points. For example, $M^* = M$ for points on the contours, $M' = \left[M^{(-)} + M^{(+)}\right]$ for points around the contours, and $M' = \left[M^{(-)} + M^{(+)} + M\right]$ for all the points. Specifically, for p_m, $p_m^{(-)}$ and $p_m^{(+)}$, we can set:

$$L \le \frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m})} \le U, \tag{23}$$

$$L \le \frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m^{(-)}})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m^{(-)}})} \le \mathrm{TH}, \tag{24}$$

$$\mathrm{TH} \le \frac{\sum_{i=1}^{Z} w_i \mathrm{WP}_i(I_{p_m^{(+)}})}{1 + \sum_{j=Z+1}^{N} w_j \mathrm{WP}_j(I_{p_m^{(+)}})} \le U, \tag{25}$$

respectively. By applying the Alternating Direction Method of Multipliers (ADMM) algorithm [30,31], this stage can quickly find a globally approximate optimal solution in high-dimensional parameter spaces and effectively prevent overfitting.

A.2. The quantitative calibration stage

Based on the initial values obtained through the qualitative calibration algorithm in Appendix A.1, we further refine the solution through iterative optimization in the quantitative calibration process. We search for the optimal perturbation $\{\delta w_n^{(k)}\}_{n=0}^{N}$ at the kth iteration to minimize the difference between the simulation values by the Wiener–Padé model and the actual measurement data. Therefore, we define the following objective function:

$$\begin{aligned}
&\min \sum_{m=1}^{M}\left(\sum_{n=1}^{N}\left[\frac{\partial R_n^{(k)}\left(x_m, y_m\right)}{\partial w^{(k)} \cdot \mathrm{SL}^{(k)}\left(x_m, y_m\right)} + \frac{\partial R_n^{(k)}\left(x_m', y_m'\right)}{\partial w^{(k)} \cdot \mathrm{SL}^{(k)}\left(x_m', y_m'\right)}\right]\delta w_n^{(k)} + \Delta \mathrm{CD}_m\right)^2 \\
&\stackrel{\text{def}}{=} \sum_{m=1}^{M}\left(\sum_{n=1}^{N} A_{mn}^{(k)}\delta w_n^{(k)} + \Delta \mathrm{CD}_m\right)^2 \\
&= \left\|\mathbf{A}^{(k)}\delta \mathbf{w}^{(k)} + \mathbf{e}^{(k)}\right\|^2,
\end{aligned} \tag{26}$$

where, M is the number of measured CDs, $\Delta \mathrm{CD}_m$ is the difference between the simulated CD and the measured CD for the mth measurement. $\mathbf{e}^{(k)} = [\Delta \mathrm{CD}_1, \Delta \mathrm{CD}_2, \ldots, \Delta \mathrm{CD}_M]^T$ is a column vector of ΔCD, $\delta \mathbf{W}^{(k)} = \left[\delta w_1^{(k)}, \delta w_2^{(k)}, \ldots, \delta w_N^{(k)}\right]^T$ is a column vector of adjustments to the Wiener–Padé coefficients, and $\mathbf{A}^{(k)}$ is a linear operator represented in a matrix form $\mathbf{A}^{(k)} = \left[A_{mn}^{(k)}\right]_{mn}$ with

$$A_{mn}^{(k)} = \frac{\partial R_n^{(k)}(I_{p_m})}{\partial w^{(k)} \cdot \mathrm{SL}^{(k)}(I_{p_m})} + \frac{\partial R_n^{(k)}(I_{p_m'})}{\partial w^{(k)} \cdot \mathrm{SL}^{(k)}(I_{p_m'})}, \tag{27}$$

where

$$\begin{aligned}
&\frac{\partial R_n^{(k)}(I_{p_m})}{\partial w^{(k)}} \\
&= \mathrm{WP}_n(I_{p_m})\left[\begin{array}{c}
\frac{\mathbf{1}_{n\le Z}}{1+\sum_{j=Z+1}^{N} w_j^{(k)}\mathrm{WP}_j(I_{p_m})} \\
+\frac{(1-\mathbf{1}_{n\le Z})\cdot\left(\sum_{i=1}^{Z} w_i^{(k)}\mathrm{WP}_i(I_{p_m})\right)}{\left(1+\sum_{j=Z+1}^{N} w_j^{(k)}\mathrm{WP}_j(I_{p_m})\right)^2}
\end{array}\right].
\end{aligned} \tag{28}$$

The spatial slopes

$$\mathrm{SL}^{(k)}(I_{p_m}) = \frac{1}{2\Delta x}\left(\frac{\sum_{i=1}^{Z} w_i^{(k)}\mathrm{WP}_i(I_{p_m+\Delta p})}{1+\sum_{j=Z+1}^{N} w_j^{(k)}\mathrm{WP}_j(I_{p_m+\Delta p})} - \frac{\sum_{i=1}^{Z} w_i^{(k)}\mathrm{WP}_i(I_{p_m-\Delta p})}{1+\sum_{j=Z+1}^{N} w_j^{(k)}\mathrm{WP}_j(I_{p_m-\Delta p})}\right) \tag{29}$$

$\mathrm{SL}^{(k)}\left(x_m', y_m'\right)$ and $\Delta \mathrm{CD}_m$ are computed based on the Wiener–Padé coefficients $\{w_n^{(k)}\}_{n=0}^{N}$ at the beginning and treated as constants for the remaining of the kth iteration. Note that the slope directions of the end points are opposite.

Based on the objective functions in Eq. (26), we can construct a constrained convex optimization function to directly solve for $\delta \mathbf{w}^{(k)}$:

$$\begin{aligned}
&\min \left\|\mathbf{A}^{(k)}\delta \mathbf{w}^{(k)} + \mathbf{e}^{(k)}\right\|^2 \\
&= \delta \mathbf{w}^{(k)^T}\left(\mathbf{A}^{(k)^T}\mathbf{A}^{(k)}\right)\delta \mathbf{w}^{(k)} + 2\mathbf{e}^{(k)^T}\mathbf{A}^{(k)}\delta \mathbf{w}^{(k)} + \mathbf{e}^{(k)^T}\mathbf{e}^{(k)} \\
&= \frac{1}{2}\delta \mathbf{w}^{(k)^T}\left(\mathbf{A}^{(k)^T}\mathbf{A}^{(k)}\right)\delta \mathbf{w}^{(k)} + \mathbf{e}^{(k)^T}\mathbf{A}^{(k)}\delta \mathbf{w}^{(k)} \\
&= \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x},
\end{aligned} \tag{30}$$

where $\mathbf{x} = \delta \mathbf{w}^{(k)}$, $\mathbf{P} = \mathbf{A}^{(k)^T}\mathbf{A}^{(k)}$, and $\mathbf{q}^T = \mathbf{e}^{(k)^T}\mathbf{A}^{(k)}$. To ensure consistency and effectiveness of the overall algorithm's constraints, we can set the constraints of the quantitative calibration algorithm to be the same as those in the qualitative calibration. By setting appropriate iteration termination conditions, and using the ADMM algorithm, this stage can quickly approximate the final optimal value by solving for $\delta \mathbf{w}$.

Data availability

Data will be made available on request.

References

[1] X. Ma, S. Zhang, Y. Pan, et al., Research and progress of computational lithography, Laser Optoelectron. Prog. 59 (9) (2022) 0922008.
[2] G. Chen, Z. Zhang, S. Li, et al., Study on deep ultraviolet computational lithography techniques, Laser & Optoelectron. Prog. 59 (9) (2022) 0922007.
[3] W.C. Huang, C.H. Lin, C.C. Kuo, C. Huang, J. Lin, J.H. Chen, R.G. Liu, Y.C. Ku, B.J. Lin, Two threshold resist models for optical proximity correction, in: Optical Microlithography XVII, 5377, SPIE, 2004, pp. 1536–1543.
[4] J. Randall, H. Gangala, A. Tritchkov, Lithography simulation with aerial image—variable threshold resist model, Microelectron. Eng. 46 (1–4) (1999) 59–63.
[5] J. Randall, K.G. Ronse, T. Marschner, A.M. Goethals, M. Ercken, Variable-threshold resist models for lithography simulation, in: Optical Microlithography XII, vol. 3679, SPIE, 1999, pp. 176–182.
[6] Y. Granik, N. Cobb, D. Medvedev, Application of CM0 resist model to OPC and verification, in: Optical Microlithography XIX, vol. 6154, SPIE, 2006, pp. 1130–1136.
[7] Y. Granik, D. Medvedev, N. Cobb, Toward standard process models for OPC, in: Optical Microlithography XX, vol. 6520, SPIE, 2007, pp. 1447–1452.
[8] Y. Granik, Transformational invariance in compact process modeling, J. Micro/Nanolithography, MEMS, MOEMS 19 (1) (2020) 013502.
[9] N. Lafferty, K. Adam, Y. Granik, A. Torres, W. Maurer, Physically based compact models for fast lithography simulation, in: Optical Microlithography XVIII, vol. 5754, SPIE, 2005, pp. 537–542.
[10] A. Chen, Y.M. Foong, D.Q. Zhang, H. Zhang, A. Chung, D. Fryer, Y. Deng, D. Medvedev, Y. Granik, Evaluation of compact models for negative-tone development layers at 20/14nm nodes, in: Optical Microlithography XXVIII, vol. 9426, SPIE, 2015, pp. 440–451.
[11] T. Kimura, T. Matsunawa, S. Mimotogi, Compact resist model using single convolution kernel, in: Optical Microlithography XXXIII, vol. 11327, SPIE, 2020, pp. 118–124.
[12] Y. Fan, C.E.R. Wu, Q. Ren, H. Song, T. Schmoeller, Improving 3D resist profile compact modeling by exploiting 3D resist physical mechanisms, in: Optical Microlithography XXVII, vol. 9052, SPIE, 2014, pp. 270–280.
[13] C. Kim, S. Lee, S. Park, N.Y. Chung, J. Kim, N. Bang, S. Lee, S. Lee, R. Boone, P. Li, et al., Machine learning techniques for OPC improvement at the sub-5 nm node, in: Extreme Ultraviolet (EUV) Lithography XI, vol. 11323, SPIE, 2020, pp. 307–320.
[14] Y. Kwon, Y. Shin, Calibration of Compact Resist Model Through CNN Training, IEEE Trans. Semicond. Manuf. (2023).
[15] C. Mu, L. Cheng, S. Zhang, H. Jiang, D.H. Wei, Y. Sun, J. Zhu, S. Liu, Efficient nonlinear resist modeling by combining and cascading quadratic Wiener systems, Opt. Laser Technol. 183 (2025) 112315.
[16] G.A. Baker Jr., J.L. Gammel, The Padé approximant, J. Math. Anal. Appl. 2 (1) (1961) 21–30.
[17] G.A. Baker Jr., The Theory and Application of the Padé Approximant Method, Technical Report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1964.
[18] J. Basdevant, The Padé approximation and its physical applications, Fortschr. Phys. 20 (5) (1972) 283–331.
[19] A.Y. Kibangou, G. Favier, M.M. Hassani, Selection of generalized orthonormal bases for second-order Volterra filters, Signal Process. 85 (12) (2005) 2371–2385.
[20] G. Palm, T. Poggio, The Volterra representation and the Wiener expansion: Validity and pitfalls, SIAM J. Appl. Math. 33 (2) (1977) 195–216.
[21] T. Ogunfunmi, Adaptive nonlinear system identification: The Volterra and Wiener model approaches, Springer Science & Business Media, 2007.
[22] K. Perev, Orthogonal approximation of Volterra series and Wiener G-functionals descriptions for nonlinear systems, Inf. Technol. Control. (2019).
[23] H. Wei, Computer simulation of photolithographic processing, 2013, US Patent 8, 532, 964.
[24] C. Mu, S. Zhang, D.H. Wei, H. Jiang, S. Liu, Wiener-Padé model for lithographic resist modeling, in: Eighth International Workshop on Advanced Patterning Solutions, vol. 13423, IWAPS 2024, SPIE, 2024, pp. 252–256.
[25] C.A. Floudas, V. Visweswaran, Quadratic optimization, Handb. Glob. Optim. (1995) 217–269.
[26] H.M. Markowitz, The optimization of a quadratic function subject to linear constraints, Nav. Res. Logist. Q. 3 (1–2) (1956) 111–133.

[27] L.T. Hoai An, An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints, Math. Program. 87 (2000) 401–426.
[28] H.K. Xu, An iterative approach to quadratic optimization, J. Optim. Theory Appl. 116 (2003) 659–678.
[29] C. Mack, Fundamental Principles of Optical Lithography: The Science of Microfabrication, John Wiley & Sons, 2007.
[30] D. Han, X. Yuan, A note on the alternating direction method of multipliers, J. Optim. Theory Appl. 155 (2012) 227–238.
[31] M. Fukushima, Application of the alternating direction method of multipliers to separable convex programming problems, Comput. Optim. Appl. 1 (1992) 93–111.