Research Article Vol. 33, No. 8 / 21 Apr 2025 / Optics Express 17280
Optics EXPRESS

Information theoretical computational lithography based on pattern density statistics

BINGYANG WANG,[1] XU MA,[1,2,*] JIAMIN LIU,[2] HAO JIANG,[2] SHIYUAN LIU,[2] AND GONZALO R. ARCE[3]

[1] *Key Laboratory of Photoelectronic Imaging Technology and System of Ministry of Education of China, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China*
[2] *State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*
[3] *Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA*
[*] *maxu@bit.edu.cn*

Abstract: Computational lithography is an important technology to improve the image resolution and fidelity of the optical lithography process. Recently, information theoretical models were introduced to explore the physical limit of image fidelity that can be achieved by different computational lithography methods. However, the existing models were derived based on a simple and idealized assumption of uniform pattern density, thus rendering a loose lower bound on the lithography imaging error. This work improves the accuracy of the information theoretical model by introducing a statistical approach of pattern density. In particular, a density classification rule (DCR) of mask and print image is established based on a number of randomly generated layout samples. The information transfer function between the mask and print image is formulated under the DCR constraint. Then, the optimal information transfer (OIT) and the theoretical limit of lithography image fidelity are derived using a numerical optimization algorithm with mask manufacturing regularization. It has been proved analytically and experimentally that our proposed model provides a much more accurate theoretical limit of lithography image fidelity than the conventional approach.

1. Introduction

Optical lithography is one of the most important processes in semiconductor manufacturing [1]. As shown in Fig. 1(a), the lithography system uses an ultraviolet light source to expose a transmissive mask carrying the integrated circuit layout, and then the aerial image of the mask pattern is conveyed onto the wafer surface through a bandlimited optical projection system [2]. After that, a series of complex processes are conducted including the photoresist development, etching, ion implanting and so on. As the critical dimension (CD) of integrated circuits step into the deep sub-wavelength realm, the lithography image of mask will be severely distorted by the optical proximity effect [3]. Thus, the computational lithography approaches must be used in advanced technology nodes to preserve the process window and the yield of high volume manufacturing [3].

Inverse lithography technique (ILT) is a leading-edge computational lithography approach that precisely compensates the lithography image distortion [4–11], where the image distortion refers to the unexpected deviation or variation compared to the target layout. As shown in Fig. 1(b), the mask pattern is represented by a binary pixelated image, where each mask pixel can be transparent (white pixel) or opaque (black pixel). In the ILT approaches, the mask pattern is inversely optimized based on the lithography process models, making the wafer image as close to the target layout as possible. Compared to the traditional mask optimization methods, ILT notably increases the degrees of optimization freedom by operating the mask pattern at pixel level, thus it may effectively enhance the lithography resolution and image fidelity.

#553893 https://doi.org/10.1364/OE.553893
Journal © 2025 Received 26 Dec 2024; revised 5 Mar 2025; accepted 28 Mar 2025; published 10 Apr 2025

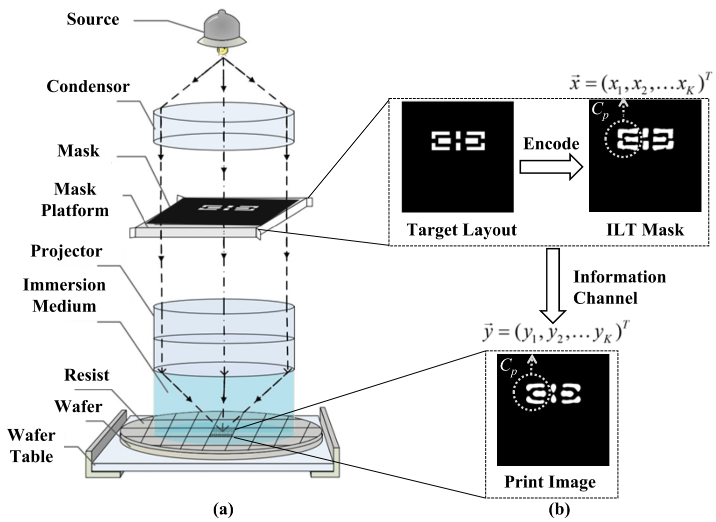

Fig. 1. (a) The sketch of DUV lithography system (for detailed descriptions of the system, please refer to Ref. [2]), and (b) the information transmission of the ILT layout from mask to wafer (revised from Fig. 1 in Ref. [12]).

An extensively used metric of lithography image fidelity is pattern error (PE) that is defined as the Frobenius norm of the difference between the actual print image on wafer and the target layout. Since the image fidelity is highly relevant to the electronic characteristics of semiconductor devices, it is natural to ask: What is the best image fidelity that ILT can reach? Indeed, it is a challenging task to find out the theoretical bound governing the attainable image fidelity, since the lithography imaging models are defined by complex nonlinear functions and the mask pattern may encompass diverse structures.

Over the past few years, researchers developed an unconventional approach by reimagining the lithography process from an information theoretical perspective [12,13]. As shown in Fig. 1(b), the lithography system can be regarded as an information channel, where the mask pattern and the wafer print image are regarded as the input signal and the output signal, respectively. Those information theoretical methods facilitate the in-depth investigation of the mask information transmission mechanisms and the theoretical limit of lithography image fidelity.

In 2017, Ma et al. pioneered an analytical information channel model that facilitates the analysis of the layout information transmission and image fidelity limit within the coherent lithography system [13]. Subsequently, this model was extended to the partially coherent lithography systems at 45 nm technology node and below, and the lower bound of PE achievable with ILT was derived [12]. Recently, the information theoretical approaches were applied to other computational lithography techniques such as the source mask co-optimization [14,15] and the phase-shifting mask optimization [16], where the information channel models were incorporated to further improve the lithography image performance.

However, the existing information theoretical models of lithography systems relied on an idealized assumption of uniform pattern density (AUPD) for simplifying the mathematical derivations. It assumes that the mask pattern and the print image on wafer can have arbitrary geometric features. Thus, for a given mask (or print image) area including a fixed number of valid pixels, the AUPD states that all kinds of distribution patterns have the equal probability. Although this simple assumption can somehow render some useful results, it does not align with the realistic situations. Practical mask patterns (or print images) always include similar geometric features as the target layouts. That means the blocked (or unblocked) pixels often cluster together, while the isolated pixels are not common. Since the AUPD is inadequate to describe the discrimination among different pixel distribution patterns, the derived theoretical limit of lithography image fidelity is not accurate enough.

This study focuses on understanding the regularities of pixel distributions in the mask and print image, and then introduces a statistical approach regarding the pattern distribution density for improving the accuracy of the information theoretical model. The innovation of this work comes from the insight that the transferred information along the lithography system is not only carried on the number of pixels, but also carried on the distribution of pixels. The concept of pixel density (PD) is introduced to measure the clustering degrees of different pixel distribution patterns. In order to calculate the value of PD, the density classification rule (DCR) is proposed based on the Monte Carlo method. In particular, we generate a large number of random pixelated patterns. For each pattern, the average distance from each valid pixel to the clustering center is calculated. Then, all of the average distances are arranged in the ascending order, and the interval of each density class is determined according to the proportional sequence. After that, given any new region on the mask (or print image), we can firstly calculate the average distance of valid pixels to the clustering center, and then map it to a certain density class and obtain the PD value.

Based on the PD and DCR, the probability mass vectors of mask pattern and print image are defined, which address not only the pixel numbers, but also the pixel distribution densities. Then, this paper derives the information transfer function and the mutual information between mask and print image under the constraint of DCR. In addition, the PE of print image is expressed as a function of the mutual information. We subsequently apply a gradient-based algorithm with manufacturing regularization [17,18] to solve for the optimal information transfer (OIT). The OIT represents the best mutual information to achieve the theoretical limit of image fidelity. The proposed methods are verified by a set of ILT experiments. We demonstrate that the image fidelity limit derived from the new model is much more accurate than the conventional model. Based on the proposed information theoretical model, we can also refine the mask pattern obtained by the traditional ILT algorithm, and further reduce the lithography image error.

2. Traditional information model based on AUPD

According to the Abbe's imaging theory, the aerial image of the lithography system is [3]:

$$\mathbf{I}(\mathbf{r}) = \sum_{\mathbf{m}} \Gamma_{\mathbf{m}} |\mathbf{h}^{\mathbf{m}}(\mathbf{r}) \otimes \mathbf{M}(\mathbf{r})|^2, \tag{1}$$

where $\mathbf{M}(\mathbf{r})$ represents the mask pattern, r is the spatial coordinate, $\otimes$ represents the convolution operation, $\mathbf{h}^{\mathbf{m}}(\mathbf{r})$ is the point spread function of the lithography system associated with the source point $\mathbf{m}$, and $\Gamma_{\mathbf{m}}$ represents the intensity of the source point. For the advanced lithography technology nodes, the imaging model in Eq. (1) can be readily modified to incorporate the mask three-dimensional (3D) effects by replacing the $\mathbf{M}$ in Eq. (1) with the diffraction near-field of the thick mask. Considering the photoresist effect, the print image on the wafer is given by:

$$\mathrm{Z} = \mathrm{sig}\{\mathrm{I},\ t_r\} = 1/\{1 + \exp[-a_r(\mathrm{I} - t_r)]\}, \tag{2}$$

where t_r is the threshold of photoresist, a_r dictates the steepness of the sigmoid function.

In [12], we proposed an information theoretical lithography model. The lithography system is regarded as an information transmission channel, and ILT is analogous to a signal encoding procedure, which facilitates the information transfer of the pre-warped mask pattern through the lithography system under the limited channel capacity. The channel capacity is essentially confined by the diffraction-limited optical system that cuts off the high frequency components of mask pattern and introduces the distortion of print image. This phenomenon is depicted by the optical proximity effect. In other words, the print image of one mask pixel is influenced by

Optics EXPRESS

several neighboring mask pixels. The influence range of optical proximity effect is determined by the coverage area of $\mathbf{h}^{\mathbf{m}}(\mathbf{r})$, which is a circular area denoted as C_P. Similarly, any pixel on the print image is correlated with several neighboring pixels covered by C_P. Consequently, the layout information is not transferred by independent pixels, but jointly transferred by a cluster of pixels from mask to wafer.

Assume the circle C_P includes K pixels, as shown in Figs. 2(a) and 2(d). Let vector $\vec{x} = (x_1, x_2, \ldots x_K)^T$ represent the mask pixels covered by C_P, and let vector $\vec{y} = (y_1, y_2, \ldots y_K)^T$ represent the K pixels on print image covered by C_P, corresponding to $\vec{x}$. Let N_x and N_y represent the numbers of one-valued pixels in $\vec{x}$ and $\vec{y}$, respectively. Assume p_x represents the probability of $\vec{x}$ containing m one-valued pixels, and q_y represents the probability of $\vec{y}$ containing n one-valued pixels, i.e.,

$$p_x = P_r\{N_x = m\}, q_y = P_r\{N_y = n\}. \tag{3}$$

In our previous works, the information entropy of $\vec{y}$ was derived based on the AUPD, which assumes that the pixels on the mask and print image have random distributions, and the probability of any distribution is equal. Specifically, for n one-valued pixels covered by C_P, there are a total of C_K^n possible distributions. Then, the probability of each distribution is $P_r\{N_y = n\}/C_K^n$. Therefore, the information entropy of $\vec{y}$ is expressed as [12]:

$$E(\vec{y}) = -\sum_{n=0}^{K}\left[\frac{P_r\{N_y = n\}}{C_K^n} \cdot \log_2\left(\frac{P_r\{N_y = n\}}{C_K^n}\right) \cdot C_K^n\right]. \tag{4}$$

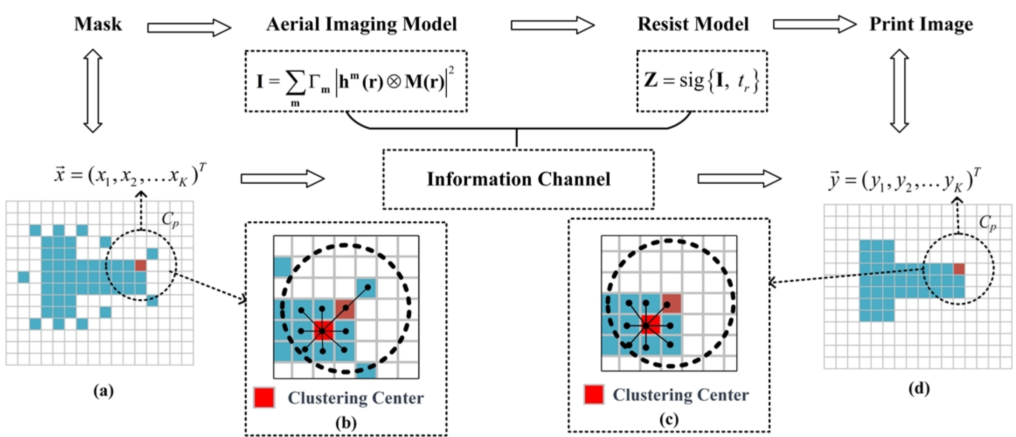

Fig. 2. The imaging model and information channel model of partially coherent lithography system, where (a) and (d) show the mask pattern and print image serving as the input and output signals; (b) and (c) illustrate the analysis method of pixel distributions based on the clustering centers.

3. Improved information model based on pattern density statistics

Based on the simplified model mentioned above, some useful conclusions were drawn in [12]. However, the AUPD is not accurate in many cases. On one hand, the pixel distributions in both mask and print image are not disorderly. Instead, the practical mask patterns (or print images) always have the similar geometric features with the target layouts. Thus, the blocked (or unblocked) pixels are likely cluster together, while the isolated and scattered pixels are unusual. In addition, considering the manufacturability of the real masks, the distribution of mask pixels should follow some regularities. On the other hand, from the information theoretical perspective,

Optics EXPRESS

the pixel distribution should not be completely random either, since the pixel distribution also carries information during the lithography imaging process. The AUPD only considers the pixel count information, but ignores the pixel arrangement information, thus leading to a loose lower-bound of PE for optical lithography system.

In order to overcome these drawbacks, this work proposes an improved information theoretical model that assigns varying probabilities for different pixel distribution patterns. In particular, a statistical approach is developed to depict the discrimination of pixel distributions. First, let's consider a circle C_P including n one-valued pixels (please refer to Fig. S1 in Section 1 of Supplement 1). The size of C_P is determined by the influence range of optical proximity effect. Then, we define a quantitative indicator called PD to measure the aggregation degrees of the pixels on mask pattern or print image. This analysis introduces a parameter E, the predefined maximum value of PD, corresponding to the E + 1 intervals for its statistical calculation. The PD metric takes the integeral values from 0 to E. The smaller PD indicates a denser distribution, while the larger PD indicates a sparser distribution. In order to calculate the value of PD, we establish the DCR based on the Monte Carlo method. The detailed descriptions of PD and DCR are provided in Section 1 of Supplement 1.

Next, let p_{ma} denote the probability of the event that $\vec{x}$ includes m one-valued pixels and the pixel density of $\vec{x}$ is $PD_x = a$. Let q_{nb} denote the probability of the event that $\vec{y}$ includes n one-valued pixels and the pixel density of $\vec{y}$ is $PD_y = b$, that is

$$p_{ma} = P_r\{N_x = m, PD_x = a\}, p_{nb} = P_r\{N_y = n, PD_y = b\}, \tag{5}$$

where $m, n = 0, 1 \ldots K$ and $a, b = 0, 1 \ldots E$. Define the vectors of probability masses for mask and print image as $\vec{p} = (p_{00}, \ldots p_{0E}, p_{10}, p_{11}, \ldots p_{KE})^T$ and $\vec{q} = (q_{00}, \ldots q_{0E}, q_{10}, q_{11}, \ldots q_{KE})^T$, respectively. Suppose $\mathrm{T} \in \mathbb{R}^{[(K+1)*(E+1)]\times[(K+1)*(E+1)]}$ is the probability transfer matrix between $\vec{p}$ and $\vec{q}$, that is

$$\vec{q} = \mathrm{T} \cdot \vec{p}, \tag{6}$$

where the element of T located in the $[(E+1)n + b + 1]$ th row and the $[(E+1)m + a + 1]$ th column is defined as $P_r\left\{N_y = n, PD_y = b \mid N_x = m, PD_x = a\right\}$, which indicates the conditional probability of $\left\{N_y = n, PD_y = b\right\}$ given $\{N_x = m, PD_x = a\}$. Similar to the method in [12], we can use several sets of ILT masks and their corresponding print images as the training samples. Based on those training samples, the matrix T can be calculated using a statistical approach. The details to calculate matrix T are given in Section 2 of Supplement 1.

Next, we proceed to derive the information entropy and mutual information between the mask and print image. Given a circular region C_P on the print image with $\{N_y = n, PD_y = b\}$, we assume that the probabilities of all distribution patterns with n one-valued pixels are the same under the condition of $PD_y = b$, and their probability can be calculated as: $P_r\{N_y = n, PD_y = b\}/\{\alpha_b C_K^n\}$, where α_b represents the proportion of the distribution patterns with $\{N_y = n, PD_y = b\}$ among all possible distribution patterns with $\{N_y = n\}$. Thus, the value of α_b can be computed based on the mapping relationship $G_n(\cdot)$ of Eq. (S3) in Section 1 of Supplement 1.

Based on the analysis above, we can calculate the mutual information between $\vec{x}$ and $\vec{y}$, denoted by $I(\vec{x};\vec{y})$, as follows:

$$I(\vec{x};\vec{y}) = -\sum_{n=0}^{K}\sum_{m=0}^{K}\sum_{b=0}^{E}\sum_{a=0}^{E}\mathbf{T}_{nbma}\cdot p_{ma}\left[\log_2\left(\sum_{u=0}^{K}\sum_{v=0}^{E}\mathbf{T}_{nbuv}\cdot p_{uv}\right) - \log_2\mathbf{T}_{nbma}\right], \tag{7}$$

where p_{ma} and p_{uv} represent the $[(E+1)m + a + 1]$th element and the $[(E+1)u + v + 1]$th element in $\vec{p}$, respectively; $\mathbf{T}_{nbma}$ is the element of T in the $[(E+1)n + b + 1]$th row and the $[(E+1)m + a + 1]$th column; and $\mathbf{T}_{nbuv}$ is the element of T in the $[(E+1)n + b + 1]$th row and the $[(E+1)u + v + 1]$th column. The detailed derivation of Eq. (7) can be found in Section 3 of Supplement 1.

Optics EXPRESS

It is shown that $I(\vec{x};\vec{y})$ can be expressed as a function solely related to the probability mass vector $\vec{p}$ of mask pattern. The mutual information represents the rate at which the information can be faithfully transmitted from mask to wafer through the lithography system.

Compared to Eq. (12) in Ref. [12], we can prove that the modified information model in this paper aligns with the actual lithography process more closely. In Supplement 1, we provide the mathematical proof in details. By considering the discrimination of pixel distribution densities, the modified model achieves a lower mutual information in contrast to the result in [12]. Consequently, the rate of error-free information transmission through lithography system is reduced, which leads to a more compact theoretical lower bound of lithography image error.

4. Relationship between PE and mutual information

In computational lithography, PE is one of the critical metrics for assessing the image fidelity of lithography systems. A commonly used definition of PE is $\mathrm{PE} = \sum_{m-1}^{N}\sum_{n-1}^{N}(\hat{\mathbf{Z}}_{mn} - \mathbf{Z}_{mn})^2$, where $\mathbf{Z}$ is the predicted print image obtained by the simulation model, $\hat{\mathbf{Z}}$ is the target layout, N is the pixel number on each edge of the layout pattern [19,20]. In this work, we utilize the information theoretical model to determine the OIT of lithography system. By solving for the OIT, we can then derive the theoretical lower bound of PE and the corresponding optimal probability distribution of mask pattern. It is noted that the proposed information theoretical model can be further extended to involve other metrics considered in the actual lithography production. Due to the length limit of the paper, this topic will be studied in the future work.

In Ref. [12], we have derived the relationship between the minimum value of PE and the mutual information, which is formulated as:

$$\mathrm{PE}_{\min} = \begin{cases} (c^* \bmod c)\cdot H_t/(2\cdot c^2) & \text{if } c^* \le \mathrm{CD} \text{ and } (c^* \bmod c)<c/2 \\ [c-(c^* \bmod c)]\cdot H_t/(2\cdot c^2) & \text{if } c^* \le \mathrm{CD} \text{ and } (c^* \bmod c)\ge c/2 \\ \min\{(c^*-\mathrm{CD})\cdot H_t/(2\cdot c^2), A_t/c^2\} & \text{if } c^*>\mathrm{CD} \end{cases}, \tag{8}$$

where c is the side length of a single pixel, $c^* = c\cdot\sqrt{K/I(\vec{x};\vec{y})}$, K is the number of pixels covered by C_P, "mod" is the congruence symbol, H_t and A_t respectively represent the perimeter and area of the target layout, CD is the critical dimension, and $\min\{\cdot,\cdot\}$ is the operation taking the minimum value. It is noted that the mutual information $I(\vec{x};\vec{y})$ is included in c^*, and thus $\mathrm{PE}_{\min}$ is a function of $I(\vec{x};\vec{y})$. The derivation of Eq. (8) can be found in [12].

Our goal is to derive the image fidelity limit of the lithography system, which is equivalent to find out the lower bound of $\mathrm{PE}_{\min}$. According to Eq. (8), the geometric interpretation of $\mathrm{PE}_{\min}$ is the minimum coverage error of covering the target layout by the macro-pixels with side length c^* [12]. It was proved that in order to reach the theoretical lower bound of PE, the side length of macro-pixel (c^*) should be the integral multiple of the side length of single pixel (c). Since $c^* = c\cdot\sqrt{K/I(\vec{x};\vec{y})}$, we should keep $I(\vec{x};\vec{y}) \approx K/\Pi^2$ [12], where Π is an integer as large as possible. Based on this condition, we can construct the following cost function:

$$F = \left[I(\vec{x};\vec{y}) - \frac{K}{\Pi^2}\right]^2. \tag{9}$$

$I(\vec{x};\vec{y})$ was described in Eq. (7). By minimizing the cost function in Eq. (9), we can obtain the optimal mask probability distribution and the OIT.

5. Theoretical limit of image fidelity and refinement of ILT solution

In this section, we will derive the OIT of lithography system and the theoretical lower bound of PE achievable by ILT. To achieve this goal, a gradient-based algorithm is used to minimize the cost function in Eq. (9), where four constraints regarding $\vec{p}$ are considered. Those constraints are proposed based on the physical properties of $\vec{p}$ and the mask design rules:

- **Constraint I:** All elements in $\vec{p}$ represent probabilities, and should satisfy $0 \le p_{ma} \le 1$.
- **Constraint II:** According to the total probability formula, the sum of all elements in $\vec{p}$ should equal to 1, i.e. $\sum_{m=0}^{K} \sum_{a=0}^{E} p_{ma} = 1$.
- **Constraint III:** The print image associated to the optimized $\vec{p}$ should be close to the target layout. Thus, the probability distribution ($\vec{q}$) of the print image should be approximate to the probability distribution ($\tilde{\vec{q}}$) of the target pattern, which can be expressed as $\tilde{\vec{q}} \approx \mathbf{T} \cdot \vec{p}$.
- **Constraint IV:** This constraint considers the mask manufacturability. It is noted that ILT often introduces numerous sub-resolution assist features (SRAFs) around the main features of mask pattern to improve the image fidelity [21]. The SRAFs with small physical sizes increase the mask complexity and pose a significant challenge in manufacturing, governed by rules such as minimum size requirement. To satisfy the rule of minimum SRAF size, one-valued pixels or zero-valued pixels should be respectively aggregated to form the single-valued clusters, while isolated pixels should be removed. In terms of PD, this implies that dense distribution patterns should occur with higher probabilities than the sparse distribution patterns. In other words, the generated SRAFs should have a greater likelihood of containing larger and continuous pixel groups, ensuring that their minimum dimensions obey the mask manufacturing rules. Thus, we design a manufacturability constraint as following:
$$p_{m0} \ge p_{m1} \ge \ldots \ge p_{mE} \text{ for } \forall m. \tag{10}$$

We replace the $I(\vec{x};\vec{y})$ in Eq. (9) with Eq. (7), and transfer the above constraints to the penalty terms in the cost function. Then, Eq. (9) can be modified as:

$$\begin{aligned} F(\vec{p}) = & \left\{ \sum_{n=0}^{K} \sum_{m=0}^{K} \sum_{b=0}^{E} \sum_{a=0}^{E} \mathbf{T}_{nbma} \cdot p_{ma} \cdot \left[\log_2 \left(\sum_{u=0}^{K} \sum_{v=0}^{E} \mathbf{T}_{nbuv} \cdot p_{uv} \right) - \log_2 \mathbf{T}_{nbma} \right] + \frac{K}{\Pi^2} \right\}^2 \\ & + \eta_1 \left(\sum_{m=0}^{K} \sum_{a=0}^{E} p_{ma} - 1 \right)^2 + \eta_2 ||\tilde{\vec{q}} - \mathbf{T} \cdot \vec{p}||_2^2 + \eta_3 \left[\sum_{m=0}^{K} \sum_{a=0}^{E} \max(0, p_{ma} - p_{ma+1}) \right], \end{aligned} \tag{11}$$

where η_1, η_2, η_3 are the weight coefficients of the penalty terms. The first, second and third penalty terms correspond to the **Constraint II**, **Constraint III**, and **Constraint IV**, respectively. Especially, in the third penalty term, if Eq. (10) holds, then $p_{ma} < p_{ma+1}$ and the penalty is 0, otherwise the penalty will be a positive number.

Considering the **Constraint I**, each element in $\vec{p}$ should be limited within [0, 1], thus we apply a parameter transform of $p_{ma} = (1 + \cos \Theta_{ma})/2$, where the probability $p_{ma} \in [0, 1]$ is relaxed to the auxiliary parameter $\Theta_{ma} \in (-\infty, +\infty)$. Considering all indexes, we have a dummy vector $\vec{\Theta} = (\Theta_{00}, \Theta_{01}, \ldots \Theta_{ma})^T$, where $m = 0, 1 \ldots K$ and $a = 0, 1 \ldots E$. The traditional ILT algorithm can effectively reduce the PE of print image. Therefore, we first use the optimized mask obtained by the traditional ILT algorithm to initialize the dummy vector $\vec{\Theta}$, so as to achieve a better initial guess of the mask probability distribution. Then, the steepest descent algorithm is used to solve the optimal $\hat{\vec{p}}$ by minimizing Eq. (11). For the detailed gradient derivation of Eq. (11), please refer to Section 4 of Supplement 1. Substitute $\hat{\vec{p}}$ into Eq. (7), we can get the OIT denoted as $\hat{I}(\vec{x}, \vec{y})$. Finally, substituting $\hat{I}(\vec{x}, \vec{y})$ into Eq. (8), the lower bound of PE can be calculated, which represents the theoretical limit of lithography image fidelity.

Similar to Ref. [12], based on the proposed information theoretical model, we can refine the optimized mask patten obtained by the traditional ILT algorithm, and further improve the

Optics EXPRESS

lithography image fidelity. Additionally, we discuss methods for SRAF generation within this framework. The details of this method are provided in Supplement 1.

6. Numerical experiments

This section provides some numerical experiments to verify the proposed model and method. In the experiments, we use a lithography system with annular illumination, whose wavelength is 193 nm, and the inner and outer partial coherence factors are 0.8 and 0.975, respectively. The numerical aperture (NA) of the lithography system is 1.25. The CD of the target layout on wafer scale is 45 nm. The photoresist threshold in Eq. (2) is set to $t_r = 0.19$.

We select two layouts, namely "Layout 1" and "Layout 2" in Fig. 3, as the testing layouts to perform the experiments. Figures 3(a) and 3(e) show the target layouts. Using the lithography imaging models in Eqs. (1) and (2), we can calculate the print images of the target layouts, as shown in Figs. 3(c) and 3(g). It is observed that the target layouts without optimization will result in inferior print images and large PEs (1128 for Layout 1 and 2808 for Layout 2).

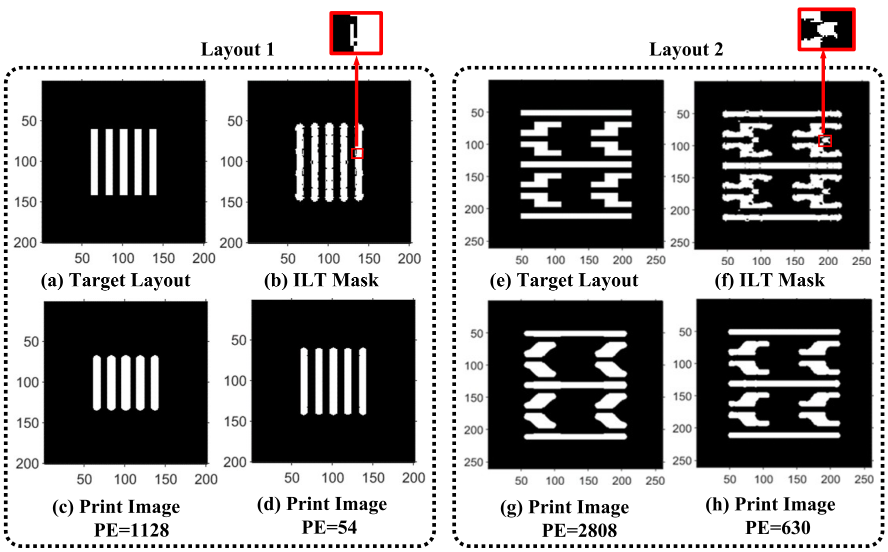

Fig. 3. Simulation results of traditional ILT algorithm for Layout 1 and Layout 2. From top to bottom, it shows the target layout, the ILT masks and their corresponding print images.

Then, we use a traditional ILT method proposed in [20] to optimize the masks. This ILT method is implemented based on the gradient-based algorithm. In order to get the best ILT solutions, we repetitively conduct the optimization simulations to explore various parameter combinations. In particular, we use the same method provided in [12] to traverse the important parameters in the algorithm. Ultimately, we choose the best ILT results with the minimum PEs. As the results, the best ILT masks are shown in Figs. 3(b) and 3(f). The corresponding print images are shown in Figs. 3(d) and 3(h). The minimum PEs achieved by the traditional ILT algorithm for these two layouts are 54 and 630, respectively.

Next, we follow the proposed method in this paper to compute the OIT and the lower bounds of PEs for the testing layouts. We take Layout 1 as an example (similarly for Layout 2) to explain this process. According to Section 5, the OIT can be calculated based on the optimal mask probability distribution $\hat{\vec{p}}$, and $\hat{\vec{p}}$ is obtained by minimizing Eq. (11) through a gradient-based algorithm. In order to accelerate the optimization convergence speed, we set the initial probability

Optics EXPRESS

$\vec{p}^0$ as the probability distribution of the ILT mask with the minimum PE. That is, for Layout 1 and Layout 2, we initialize the $\vec{p}^0$ according to the ILT masks in Fig. 3(b) and 3(f), respectively.

In Section 1 of Supplement 1, we describe how to use the Monte Carlo method to determine the DCR, and then calculate the PD. Combining the counts of one-valued pixels and the PD information for all C_p, the initial mask probability distribution $\vec{p}^0$ can be determined based on Eq. (5). Figures 4 (a) and 4(b) show the $\vec{p}^0$'s for Layout 1 and Layout 2, respectively. In these figures, the x-axis represents the sequential index of each component in $\vec{p}^0$, while the y-axis (in logarithmic scale) represents the probability value corresponding to each index.

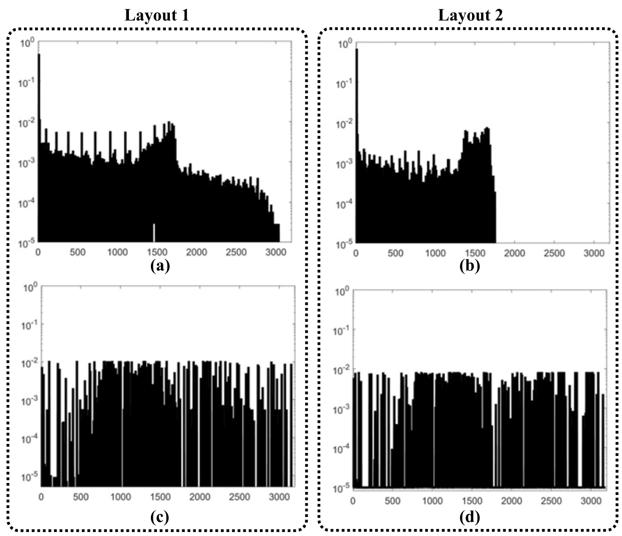

Fig. 4. The initiasl mask probability distributions $\vec{p}^0$ for (a) Layout 1 and (b) Layout 2, and the optimal mask probability distributions $\hat{\vec{p}}$ for (c) Layout 1 and (d) Layout 2.

Given the matrix $\mathbf{T}$ and $\vec{p}^0$, we use the steepest descent algorithm to minimize the cost function in Eq. (11), and obtain the optimal mask probability distribution $\hat{\vec{p}}$. The optimization parameters are set as follows. In Eq. (11), $\eta_1 = 1$, $\eta_2 = 150$ and $\eta_3 = 0.01$. The step size of the steepest descent algorithm is 0.1, and the iteration number is 100. Those parameters are determined through multiple experiments to achieve the promising convergence results.

Figures 4(c) and 4(d) display the optimal distributions $\hat{\vec{p}}$ for Layout 1 and Layout 2, respectively. Substituting $\hat{\vec{p}}$ into Eq. (7) and Eq. (8), we can calculate the OIT and the lower bound of PE. Table 1 lists the lower bounds of PEs, OITs and the minimum PEs that can be achieved by ILT methods for both Layout 1 and Layout 2. The second column and the third column respectively present the lower bounds of PEs and OITs calculated by the proposed method in this paper. The fourth column and the fifth column respectively present the lower bounds of PEs and OITs calculated by the traditional information theoretical method in [12], where the simple assumption of AUPD is used. The sixth column shows the minimum PEs achievable by the traditional ILT algorithm for both layouts, which are consistent with the simulation results in Fig. 4.

According to Table 1, we find that the proposed method in this paper provides much better theoretical limit of image fidelity than the traditional method in [12]. The evidence is obvious that the lower bounds of PEs calculated by the proposed method are closer to the actual minimum PEs obtained by the real ILT algorithm. This demonstrates the merit of this work, since the involvement of pattern density statistics, instead of using simple AUPD assumption, will effectively improve the accuracy of the information theoretical model. Moreover, the theoretical lower bounds of PEs are always less than the minimum PEs of ILT algorithm. This indicates that

Table 1. Theoretical lower bounds of PEs, OITs, and minimum PEs obtained by traditional and refined ILT methods

Layout	Lower bound of PE for proposed method	OIT for proposed method	Lower bound of PE for method in [12]	OIT for method in [12]	Minimum PE for traditional ILT	Improved PE for refined ILT in [12]
Layout 1	26.18	3.8635	3.05	4.5982	54	46
Layout 2	174.22	3.7881	17.50	4.5900	630	599

no matter how to optimize the masks, it cannot reach better print images over the theoretical limits. This once again demonstrates the correctness of the proposed information theoretical model.

Next, we consider how to improve the results of traditional ILT algorithm using the proposed information theoretical model. Comparing the initial and optimal mask probability distributions in Fig. 4, we can get the following observations. First, the distribution of $\vec{p}^0$ is less uniform than that of $\hat{\vec{p}}$. Especially, when the number of one-valued pixels is 0 (corresponding to the opaque regions on the mask), $\vec{p}^0$ has the largest probability. When the number of one-valued pixels is large, $\vec{p}^0$ has very small or even zero probability. In contrast, the distribution of $\hat{\vec{p}}$ is much more uniform. Therefore, to make the mask distribution closer to the optimal one, we need to add one-valued pixels on the mask, so that the probability of opaque regions is reduced, and the probability of opening regions is increased.

The analysis mentioned above indicates that the solution of the traditional ILT method can be further improved by inserting additional SRAFs on the mask. The SRAFs refer to the small opening areas that modulate the geometrical environments of the mask pattern, which can improve lithography image fidelity [22]. Additionally, the inserted SRAFs should make the mask have more uniform probability distribution, similar to Fig. 4(c) and Fig. 4(d). Therefore, the positions of SRAFs should be carefully arranged to maintain this uniformity across different mask regions. Based on these findings, we can refine the solutions of the traditional ILT method to push the image fidelity towards the theoretical limit.

Here, we apply the method proposed in [12] to refine the traditional ILT masks, where the information theoretical model is used to insert SRAFs in the mask pattern. The details of this refined ILT method can be found in [12] and Supplement 1. The seventh column in Table 1 presents the improved PEs for both layouts obtained by the refined ILT method, which are lower than the PEs directly obtained by the traditional ILT method. It shows that the proposed information theoretical model can help to further optimize the mask patterns, thus improving the image fidelity of lithography system. In terms of computational efficiency, incorporating the DCR will increase the computational time compared to the traditional AUPD model. It is noted that different modules in the proposed model can be handled and calculated independently, making them well-suited for the GPU-based parallel computing. By leveraging the GPU acceleration, the computational time can be significantly reduced for the simulation of large-scale integrated circuit designs. However, this topic is out of the scope of this paper, and will be studied in our future work.

7. Conclusion

This paper proposed an improved information theoretical model for advanced lithography system, and explored the theoretical limit of image fidelity for computational lithography approach. By introducing the DCR, we derived the mutual information between mask and print image, thus successfully considering the extra information transfer capacity of pixel distributions. Subsequently, the relationship between mutual information and lithography image error was discussed, and a gradient-based algorithm was used to pursue the OIT and the lower bound of PE achievable by the ILT methods. Both mathematical proof and numerical experiments were provided to verify the proposed information theoretical approaches. It shows that the proposed methods in this work can achieve much more accurate theoretical limit of lithography image fidelity compared to the existing methods. In future work, we will further improve this information theoretical model to consider the systematical errors and random variations in the real lithography process.

Funding. National Natural Science Foundation of China (62374016, 52130504); State Key Lab of Digital Manufacturing Equipment and Technology (DMETKF2022011).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography* (SPIE, 2001).
2. F. Schellenberg, "A little light magic," IEEE Spectrum. **40**(9), 34–39 (2003).
3. X. Ma and G. R. Arce, *Computational Lithography*, 1st ed. (John Wiley and Sons, 2010).
4. Y. Wei, Y. Su, L. Dong, *et al.*, *Computational Lithography and Layout Optimization* (Publishing House of Electronics Industry, China, 2021).
5. L. Pang, "Inverse lithography technology: 30 years from concept to practical, full-chip reality," J. Micro/Nanolith. MEMS, MOEMS **20**(3), 030901 (2021).
6. B. Jiang, L. Liu, Y. Ma, *et al.*, "Neural-ILT 2.0: Migrating ILT to Domain-Specific and Multitask-Enabled Neural Network," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **41**(8), 2671–2684 (2022).
7. X. Ma, Z. Wang, L. Yanqiu, *et al.*, "Fast optical proximity correction method based on nonlinear compressive sensing," Opt. Express **26**(11), 14479–14498 (2018).
8. Y. Shen, N. Jia, N. Wong, *et al.*, "Robust level-set-based inverse lithography," Opt. Express **19**(6), 5511–5521 (2011).
9. W. Lv, E. Y. Lam, H. Wei, *et al.*, "Cascadic multigrid algorithm for robust inverse mask synthesis in optical lithography," J. Micro/Nanolithogr., MEMS, MOEMS **13**(2), 023003 (2014).
10. W. Lv, S. Liu, Q. Xia, *et al.*, "Level-set-based inverse lithography for mask synthesis using the conjugate gradient and an optimal time step," J. Vac. Sci. Technol., B:Nanotechnol. Microelectron.:Mater., Process., Meas., Phenom. **31**(4), 41605 (2013).
11. K. Hooker, B. Kuechler, A. Kazarian, *et al.*, "ILT optimization of EUV masks for sub-7 nm lithography," Proc. SPIE **10046**, 1044604 (2017).
12. Z. Wang, X. Ma, G. R. Arce, *et al.*, "Information theoretical approaches in computational lithography," Opt. Express **26**(13), 16736–16751 (2018).
13. X. Ma, H. Zhang, Z. Wang, *et al.*, "Information theoretical aspects in coherent optical lithography systems," Opt. Express **25**(23), 29043–29057 (2017).
14. X. Ma, Y. Pan, S. Zhang, *et al.*, "Informational lithography approach based on source and mask optimization," IEEE Trans. Comput. Imag. **7**, 32–42 (2021).
15. Y. Pan, X. Ma, S. Zhang, *et al.*, "Efficient informatics-based source and mask optimization for optical lithography," Appl. Opt. **60**(27), 8307–8315 (2021).
16. X. Ma and Y. Pan, "Informatics-based computational lithography for phase-shifting mask optimization," Opt. Express **30**(12), 21282–21294 (2022).
17. X. Ma, "Resolution enhancement optimization methods in optical lithography with improved manufacturability," J. Micro/Nanolithogr., MEMS, MOEMS **10**(2), 023009 (2011).
18. F. Liu and X. Shi, "An efficient mask optimization method based on homotopy continuation technique," in *Proc. IEEE/ACM Design Automation and Test in Europe (DATE)*, pp. 1–6 (2011).
19. W. Lv, Q. Xia, and S. Liu, "Mask-filtering-based inverse lithography," J. Micro/Nanolithogr., MEMS, MOEMS **12**(4), 043003 (2013).
20. X. Ma and G. R. Arce, "Binary mask optimization for inverse lithography with partially coherent illumination," J. Opt. Soc. Am. A **25**(12), 2960–2970 (2008).
21. Z. Zhang, S. Li, X. Wang, *et al.*, "Fast heuristic-based source mask optimization for EUV lithography using dual edge evolution and partial sampling," Opt. Express **29**(14), 22778–22795 (2021).
22. C. Chen, D. Kim, D. H. Park, *et al.*, "Improving inspectability with KLA-Tencor TeraScan Thin Line De-sense," Proc. SPIE **6730**, 67303O (2007).