Adaptive gradient-based source and mask co-optimization with process awareness

Yijiang Shen (沈逸江)[1,*], Fei Peng (彭 飞)[1], Xiaoyan Huang (黄小燕)[1], and Zhenrong Zhang (张振荣)[2]

[1]*School of Automation, Guangdong University of Technology, Mega Education Center South, Guangzhou 510006, China*

[2]*Guangxi Key Laboratory of Multimedia Communications and Network Technology, School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China*

Corresponding author: yjshen@gdut.edu.cn

Received April 24, 2019; accepted August 6, 2019; posted online December 2, 2019

We develop a source and mask co-optimization framework incorporating the minimization of edge placement error (EPE) and process variability band (PV Band) into the cost function to compensate simultaneously for the image distortion and the increasingly pronounced lithographic process conditions. Explicit differentiable functions of the EPE and the PV Band are presented, and adaptive gradient methods are applied to break symmetry to escape suboptimal local minima. Dependence on the initial mask conditions is also investigated. Simulation results demonstrate the efficacy of the proposed source and mask optimization approach in pattern fidelity improvement, process robustness enhancement, and almost unaffected performance with random initial masks.

OCIS codes: 110.5220, 100.0100, 110.1758.

doi: 10.3788/COL201917.121102.

Optical microlithography is increasingly challenging with the ever growing integration intensity of semiconductor devices in the sub-22-nm technology node and low k_1 regime. To this end, resolution enhancement techniques (RETs)[1,2] become essential for printing a good quality wafer image including modified illumination schemes, rule-based and model-based optical proximity correction (OPC)[3]. Moving beyond model-based OPC, the inverse lithography technique (ILT)[4,5] inverts the imaging model and attempts to directly synthesize the optimized mask pattern. With the development of pixelated sources[6], source and mask optimization (SMO) becomes an integral part of ILT to improve the imaging performance by expanding the solution space of the source and mask with the joint optimization of the illumination and mask shapes[7,8].

Various computational strategies including pixelated patterns[7–10], pupil and mask topology compensation[11], Zernike source representations[12], wave front modulation[13–15], and compressive sensing[16,17] are incorporated into the SMO framework, which is readily solved by gradient-based methods[18–21]. Special attentions have been paid to dose sensitivity[18], defocus[19], and dose-focus matrix[22]. However, process variability band (PV Band), one important criterion for measuring process manufacturability indicating the physical representation of the layout sensitivity to process variations, is too complicated to be explicitly incorporated into the cost functions. Similarly, edge placement error (EPE) which evaluates the printed image contour under nominal conditions, is often excluded because of lack of differentiable formulations.

Gao *et al.*[23] developed objective formulations of EPE and PV Band with a scalar lithographic imaging model. Practically, selections of the step size in gradient-based methods generally face the dilemma where too small step-size subjects slow convergence and too large step-size fluctuation is around the minimal or even divergence. Besides, for sparse source and mask patterns with very different feature frequencies, updating them to the same extent is not appropriate where large updates should be performed for rarely occurring features. Accordingly, adaptive gradient method such as AdaGrad performs smaller updates for frequently occurring features and large updates for infrequent ones, and adaptive moment estimation (Adam) computes adaptive learning rates by keeping exponentially decaying averages of past square gradients and momentum. Therefore stability and the ability to escape suboptimal minimals are duly detected in the updating process.

This Letter focuses on the application of adaptive gradient methods including Adam and AdaGrad to lithographic SMO, which simultaneously considers pattern design in terms of pattern error (PE), EPE, and process window. We present explicit formulations of differentiable functions for EPE and PV Band, whose closed-form gradients are subsequently developed with vector imaging formation. Source patterns, where usually more sparsity is observed, and mask patterns are updated with AdaGrad and Adam methods, respectively. We also investigate the stability of the optimization process and the ability to escape suboptimal local minima when random initial masks are applied. Simulations show that the proposed SMO approach improves pattern fidelity and the process window with enhanced stability and unaffected initial condition performance.

The wafer imaging process $\mathcal{T}$ can be divided into two function blocks, namely the projection optics effects (coupling image formation) in Fig. 1 and resist effects.

 © 2019 Chinese Optics Letters

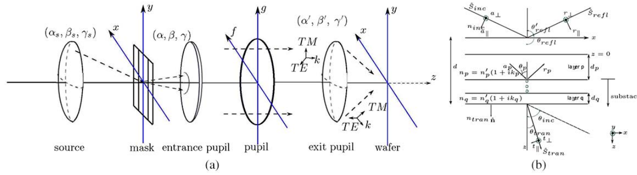

Fig. 1. (a) Schematic of forward lithography. (b) Reflection from and transmission through a stratified medium.

For a point source (α_s, β_s) emanating a polarized electric field, the coupling image $\mathbf{I}_c$ can be described as[3,20]

$$\mathbf{I}_c = \frac{1}{J_{\text{sum}}} \sum_{\alpha_s,\beta_s} \mathbf{J}(\alpha_s,\beta_s) \sum_{p=x,y,z} \|\mathbf{H}_p(\alpha_s,\beta_s) \otimes \mathbf{B}(\alpha_s,\beta_s) \odot \mathbf{M}\|^2, \tag{1}$$

where $\mathbf{J}$ is an $N_s \times N_s$ scalar matrix representing the source pattern distribution, J_{sum} is the sum of nonzero source intensities, $\mathbf{H}_p(\alpha_s,\beta_s), p = x, y, z$ are referred to as the equivalent filters, $\mathbf{B}(\alpha_s,\beta_s)$ is the diffraction matrix to approximate the mask near-field, and $\|\cdot\|^2$ means taking a pixel-wise square of amplitude. The resist effect can be approximated using a logarithmic sigmoid function $\text{sig}(x) = \frac{1}{1+e^{-a(x-t_r)}}$ with a being the steepness of the sigmoid function and t_r being the threshold. Therefore, the wafer imaging formation $\mathcal{T}(\cdot)$ is described as $\mathbf{I} = \mathcal{T}(\mathbf{J},\mathbf{M}) = \text{sig}(\mathbf{I}_c)$.

Given a target pattern $\mathbf{I}_0 \in \mathbb{R}^{N\times N}$, the goal of the SMO is to find the optimal source $\hat{\mathbf{J}} \in \mathbb{R}^{N_s\times N_s}$ and mask pattern $\hat{\mathbf{M}} \in \mathbb{R}^{N\times N}$, which minimize the measured dissimilarity or "score $(\mathcal{S})$" between $\mathcal{T}(\cdot)$ and $\mathbf{I}_0$, namely,

$$(\hat{\mathbf{J}}, \hat{\mathbf{M}}) = \min_{\mathbf{J}\in\mathbb{R}^{N_s\times N_s}} \min_{\mathbf{M}\in\mathbb{R}^{N\times N}} \mathcal{S}\{\mathcal{T}(\mathbf{J},\mathbf{M}), \mathbf{I}_0\}, \tag{2}$$

in which the formula of $\mathcal{S}$ in this work is defined as

$$\mathcal{S} = \gamma_{\text{pe}}\mathcal{S}_{\text{pe}}\{\mathbf{I}_0,\mathbf{I}\} + \gamma_{\text{epe}}\mathcal{S}_{\text{epe}}\{\mathbf{I}_0,\mathbf{I}\} + \gamma_{\text{pv}}\mathcal{S}_{\text{pv}}\{\mathbf{I}_0,\mathbf{I}\}, \tag{3}$$

where $\mathcal{S}_{\text{pe}}$, $\mathcal{S}_{\text{epe}}$, and $\mathcal{S}_{\text{pv}}$ ensure pattern fidelity, minimize the EPE and the PV Band, respectively, and are weighted by predefined weight parameter $\gamma = \{\gamma_{\text{pe}}, \gamma_{\text{epe}}, \gamma_{\text{pv}}\}$. Parametric transformations $\text{M} = 0.5 \times [1 + \cos(\omega)]$ and $\text{J} = 0.5 \times [1 + \cos(\theta)]$, with $\theta \in \mathbb{R}^{N_s\times N_s}$ and $\omega \in \mathbb{R}^{N\times N}$, are applied to reduce the binary-constrained optimization problems to unconstrained ones in the updating procedure.

$\mathcal{S}_{\text{pe}}$ measures the sum of mismatches between $\mathbf{I}$ and the desired one $\mathbf{I}_0$ over all locations. For mathematical convenience, the square of the ℓ_2 norm is frequently practiced in SMO, leading to the minimization of

$$\mathcal{S}_{\text{pe}}(\mathbf{J},\mathbf{M}) = 0.5 \times \|\mathcal{T}(\mathbf{J},\mathbf{M}) - \mathbf{I}_0\|^2. \tag{4}$$

The gradients of $\mathcal{S}_{\text{pe}}$ with respect to ω and θ are

$$\frac{\partial \mathcal{S}_{\text{pe}}}{\partial \omega} = -\frac{a\sin\omega}{2J_{\text{sum}}} \odot \sum_{\alpha_s,\beta_s} \mathbf{J} \sum_{p=x,y,z} \times \text{Real}\Big\{\mathbf{B}^* \odot \Big\{\mathbf{H}_p^{*\circ} \otimes \Big[\mathbf{E}_p \odot (\mathbf{I}-\mathbf{I}_0) \odot \mathbf{I} \odot (\mathbf{1}-\mathbf{I})\Big]\Big\}\Big\}, \tag{5}$$

$$\frac{\partial \mathcal{S}_{\text{pe}}}{\partial \theta} = -\frac{\sin\omega}{2} \sum_{x,y} \left[a \cdot (\mathbf{I}-\mathbf{I}_0) \odot \mathbf{I} \odot (\mathbf{1}-\mathbf{I}) \odot \frac{\sum_{p=x,y,z} |\mathbf{E}_p|^2 - \mathbf{I}_c}{J_{\text{sum}}} \right], \tag{6}$$

where $\odot$ is entry-by-entry multiplication, $*$ is the conjugate operation, $\circ$ rotates the matrix in the argument by 180° in both the horizontal and vertical directions, $\otimes$ is the convolution operation, $\mathbf{1} \in \mathbb{R}^{N\times N}$ is the all-ones matrix, and $\mathbf{E}_p(\alpha_s,\beta_s) = \mathbf{H}_p(\alpha_s,\beta_s) \otimes \mathbf{B}(\alpha_s,\beta_s) \odot \mathbf{M}$.

$\mathcal{S}_{\text{epe}}$ measures the geometrical distance of the image contour between $\mathbf{I}_0$ and $\mathbf{I}$. However, lack of analytic formulation of a differentiable $\mathcal{S}_{\text{epe}}$ often complicates the explicit incorporation of EPE minimization. To this end, we formulate EPE as illustrated in Fig. 2(a) to include image difference D_{sum} in the horizontal and vertical inner image and outer image edges from sampled points on horizontal edges (HS) and vertical edges (VS). EPE violation is detected to be one when $D_{\text{sum}} \geq t_e$, with t_e being a predefined threshold and zero otherwise. D_{sum} is computed for samples on vertical and horizontal edges within LH and LV, horizontal and vertical tolerable EPE segments depicted in Fig. 2(b). LH and LV are calculated according to the pattern edge set (PES) in Fig. 2(c) enwrapping the target pattern edge (TPE) in Fig. 2(d), under possible exposure latitude[1] describing tolerable target pattern linewidth. Subsequently, D_{sum} is calculated as

$$D_{\text{sum}}(i,j) = \begin{cases} \sum_{k=j-\frac{\text{LV}}{2}}^{j+\frac{\text{LV}}{2}} \mathcal{S}_{\text{pe}}(i,k) & \text{if } (i,j) \in \text{HS} \\ \sum_{k=j-\frac{\text{LH}}{2}}^{j+\frac{\text{LH}}{2}} \mathcal{S}_{\text{pe}}(k,j) & \text{if } (i,j) \in \text{VS} \end{cases}, \tag{7}$$

where $\mathcal{S}_{\text{pe}}(i,k)$ is the image difference between sampled points on HS with horizontal coordinate i and points in LH with horizontal coordinate i and vertical coordinate k. With $\mathcal{S}_{\text{pe}}$ defined in Eq. (4), $\mathcal{S}_{\text{pe}}(k,j)$ is similarly defined. $\mathcal{S}_{\text{epe}}$ is defined to be the summation of EPE violations (EPEVs) for all samples on HS and VS as

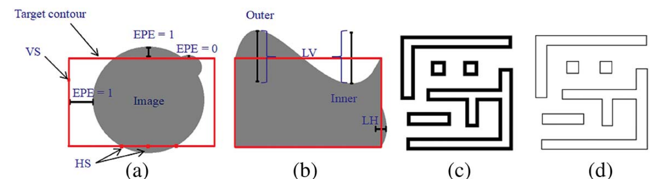

Fig. 2. (a) EPE measurement illustration. (b) Numerical superposition region. (c) Pattern edge set (PES). (d) Edges of target pattern $\mathbf{I}_{02}$ in Fig. 4(c).

$$\mathcal{S}_{\text{epe}} = \sum_{(\text{HS},\text{VS})\in\text{TPE}} \text{EPEVs}. \tag{8}$$

For $\mathcal{S}_{\text{epe}}$'s differentiability, another sigmoid function $\text{sig}_e(x) = \frac{1}{1+e^{-a_e(x-t_e)}}$ is applied to D_{sum}, removing the binary-value constraints on EPE with a_e being the steepness and t_e being the threshold of sig_e. Consequently, the gradients of $\mathcal{S}_{\text{epe}}$ with respect to ω and θ are calculated as

$$\frac{\partial \mathcal{S}_{\text{epe}}}{\partial \phi} = \sum_{(\text{HS},\text{VS})\in\text{TPE}} \sum_{(i,j)\in\text{HS or VS}} \frac{\partial \text{sig}_e(D_{\text{sum}(i,j)})}{\partial \phi}, \tag{9}$$

with $\phi = \omega$ or θ and

$$\frac{\partial \text{sig}_e(D_{\text{sum}})}{\partial \phi} = a_e \text{sig}_e(D_{\text{sum}})[1 - \text{sig}_e(D_{\text{sum}})] \sum_{k=j-\frac{\text{LV}}{2}}^{j+\frac{\text{LV}}{2}} \frac{\partial \mathcal{S}_{\text{pe}}}{\partial \phi}, \tag{10}$$

in which $\frac{\partial \mathcal{S}_{\text{pe}}}{\partial \phi}$ is defined in Eqs. (5) and (6).

PV Band is a set of edges between the fix-printability areas (FPAs) and non-printability areas (NPAs) under possible process conditions, representing the robustness of process manufacturing. As illustrated in Fig. 3, the formulation of the PV Band in Fig. 3(d) requires a series of Boolean operations to extract the edge placement through all possible printed images from Figs. 3(a) to 3(c), which are extremely cumbersome and difficult to calculate. The red boxes present extracted edges of the target contact pattern, and the gray areas are the printed patterns with the extracted pattern edges in blue. $\mathcal{S}_{\text{pv}}$ in Eq. (3) is defined as

$$\mathcal{S}_{\text{pv}} = \left(\mathbf{I}_1 \cup \mathbf{I}_2 \cup \ldots \cup \mathbf{I}_{N_p-1} \cup \mathbf{I}_{N_p}\right) \backslash \left(\mathbf{I}_1 \cap \mathbf{I}_2 \cap \ldots \cap \mathbf{I}_{N_p-1} \cap \mathbf{I}_{N_p}\right), \tag{11}$$

where $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_{N_p-1}, \mathbf{I}_{N_p}$ are printed images under N_p process conditions, $\cup$ and $\cap$ are union and intersection operations, and the operation $\backslash$ denotes the complement set of FPA in $(1 - \text{NPA})$. Noting $\text{FPA} \subset \mathbf{I}_k$, $k = 1, 2, \ldots, N_p$,

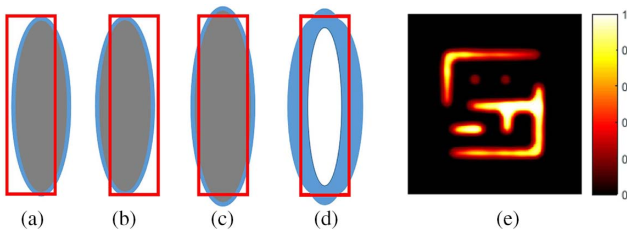

Fig. 3. PV Band demonstration. (a)–(c) Printed images under different process conditions. (d) Computed PV Band. (e) PV Band of the printed images with $\mathbf{I}_{02}$ in Fig. 4(c) illuminated by the annular source in Fig. 4(a).

$$\mathcal{S}_{\text{pv}} = (\mathbf{I}_1 \backslash \text{FPA}) \cup (\mathbf{I}_2 \backslash \text{FPA}) \cup \ldots \cup (\mathbf{I}_{N_p} \backslash \text{FPA}). \tag{12}$$

Assuming the edge of the printed pattern is close enough to the desired printed pattern edge when $\mathcal{S}_{\text{epe}}$ is incorporated in the cost function and replacing FPA with the target pattern $\mathbf{I}_0$, $\mathcal{S}_{\text{pv}}$ is reduced to the average of the summation of the ℓ_2 norm of image differences to give

$$\mathcal{S}_{\text{pv}} = \frac{1}{N_p} \sum_{k=1}^{N_p} \|\mathbf{I}_k - \mathbf{I}_0\|^2 = \frac{1}{N_p} \sum_{k=1}^{N_p} \mathcal{S}_{\text{pe}_k}, \tag{13}$$

with $\mathcal{S}_{\text{pe}_k}$ being the image difference under the kth process condition with $\mathcal{S}_{\text{pe}}$ defined in Eq. (3). Figure 3(e) shows the PV Band calculated using $\text{FPA} = 0$ and $\mathbf{M} = \mathbf{I}_0$. Therefore, the gradients of $\mathcal{S}_{\text{pv}}$ with respect to ω and θ can be routinely calculated according to Eqs. (4) and (5) as

$$\frac{\partial \mathcal{S}_{\text{pv}}}{\partial \phi} = \frac{1}{N_p} \sum_{k=1}^{N_p} \frac{\partial \mathcal{S}_{\text{pe}_k}}{\partial \phi}. \tag{14}$$

Gradient-based searching such as steepest gradient descent (SGD) has been a preferred algorithm for the minimization of $\mathcal{S}$ in Eq. (3). However, suffering from the sensitivity to step-size η, SGD is often subject to running into unwanted local minimal with small η and divergence if η is too big. Moreover, the sparsity of $\frac{\partial \mathcal{S}}{\partial \phi}$, $\phi = \omega$ or θ aggregates the dilemma of η selection. Adam method combines the merits of AdaGrad and RMSPro methods, which works well with sparse gradients and naturally performs adaptive adjustments of η. Therefore, in this Letter, AdaGrad and Adam methods are applied to updating the source and mask patterns θ or ω. In the Adam method, $\phi = \omega$ or θ at time-step $t + 1$ is updated as

$$\phi_{t+1} = \phi_t - \eta \Delta \phi_t = \phi_t - \eta \cdot \hat{m}_t / \left(\sqrt{\hat{v}_t} + \epsilon\right), \tag{15}$$

where $\epsilon = 10^{-8}$ is the smoothing term to avoid division by zero, and $\hat{m}_t = m_t/(1 - \beta_1^t)$ and $\hat{v}_t = v_t/(1 - \beta_2^t)$ are the bias-corrected moment estimate of first moment $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ and second moment $v_t = v_{t-1} \cdot \beta_2 + (1 - \beta_2) \cdot g_t^2$, respectively, with $g_t = \frac{\partial \mathcal{S}}{\partial \phi}$, $g_t^2 = g_t \cdot g_t$, and β_1, β_2 being the decay rates.

Assuming after initial optimization (IO) of ϕ which accumulates m_t and v_t, ϕ reaches a local minimum point at $t = t_1$, where SGD cannot break symmetry, with $g_{t_1} \approx 0$ and $m_t, v_t \gg g_t$, $\Delta \phi_t$ at $t = t_2$ can be calculated as

$$|\Delta \phi_t| = \frac{\prod_{t=t_1}^{t_2} \left|\beta_1 / \left(1 - \beta_1^{t+1}\right)\right| \cdot |m_{t_2}|}{\prod_{t=t_1}^{t_2} \left|\beta_2 / \left(1 - \beta_2^{t+1}\right)\right|^{1/2} \cdot |v_{t_2}|^{1/2}} = \prod_{t=t_1}^{t_2} |v_t| \cdot |\Delta \phi_{t_2}|, \tag{16}$$

in which β_1^t, β_2^t and v_t are regarded as the attenuation factors of m_{t_2}, v_{t_2}. It is therefore concluded that after

the IO procedure of accumulating m_t and v_t, the attenuation factors gradually decrease m_t and v_t small enough to be close to zero, namely as the first-phase optimization (FPO).

Subsequently, we investigate the absolute value of $\Delta\boldsymbol{\phi}_t$ at the end of (FPO) $t = t_2$ as

$$\begin{aligned}|\Delta\boldsymbol{\phi}_t| &= \frac{|[\beta_1\cdot m_t + (1-\beta_1)\cdot g_t]/(1-\beta_1^t)|}{\{[\beta_2\cdot v_t + (1-\beta_2)\cdot g_t^2]/(1-\beta_2^t)\}^{1/2}+\epsilon} \\ &= \frac{|\rho m_t|\cdot|g_t|}{|\rho v_t\cdot g_t^2|^{1/2}+\epsilon},\end{aligned} \tag{17}$$

where $\rho m_t = (1-\beta_1)/(1-\beta_1^t)$ and $\rho v_t = (1-\beta_2)/(1-\beta_2^t)$ are amplification factors with respect to g_t and g_t^2. With m_0, v_0, and g_0 close to 0, $|\Delta\boldsymbol{\phi}_t| \approx 0.5$, taking the smoothing term $\epsilon = 10^{-8}$ into account: at $t = t_2 + 1$, if g_1 is close to 0, $m_1 \approx 0$ and $v_1 \approx 0$, the iteration will act similarly to the iteration at $t = t_2$ and similarly for the following iterations until g_t deviates significantly from zero. We name the above procedure the second-phase optimization (SPO), at the end of which $|\Delta\boldsymbol{\phi}_t|$ is big enough to drive the updating of $\boldsymbol{\phi}$ out of the SPO entering IO to escape the local minimum point.

Numerical simulations are performed on a lithography imaging system with wavelength $\lambda = 193$ nm, NA = 1.35, spatial resolution $\Delta x = \Delta y = 4$ nm/pixel, $a = 80$, and $t_r = 0.25$ being the steepness and the threshold of the sigmoid function. The system is initially illuminated by an annular source with $\sigma_{\text{in}} = 0.6$ and $\sigma_{\text{out}} = 0.9$ in Fig. 4(a), with target patterns $\mathbf{I}_{01}$, $\mathbf{I}_{02}$ in Figs. 4(b) and 4(c). The ranges of process conditions including dose, defocus, and linewidth tolerance are ±2, ±50 nm, and ±10%, respectively. $\mathbf{H}_p(\alpha_s, \beta_s)$ is calculated according to the parameters of the wafer stack given in Table 1. The corresponding $\mathbf{I}$, EPE, and PV Band images when printing $\mathbf{I}_{01}$ and $\mathbf{I}_{02}$ on the wafer illuminated by $\mathbf{J}_0$ are given in Figs. 5(a)–5(c) and Figs. 5(d)–5(f), respectively. Severe distortions are observed exhibiting $\mathcal{S}_{\text{pe}}$ 4494 and 5193, $\mathcal{S}_{\text{epe}}$ 1158 and 1512 with respect to $\mathbf{I}_{01}$ and $\mathbf{I}_{02}$, respectively. Violations of linewidth tolerance are also detected with $\mathcal{S}_{\text{pv}}$ 2347 and 3965 in Figs. 5(c) and 5(f), which has to be compensated for by radical computational techniques. When updating $\boldsymbol{\phi} = \boldsymbol{\omega}$ or θ at time-step $t = t + 1$ using the SGD method with

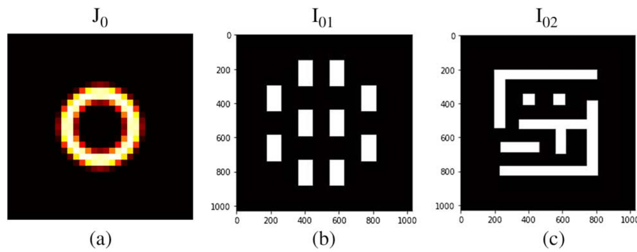

Fig. 4. (a) Annular source $\mathbf{J}_0$ with $\sigma_{\text{in}} = 0.6$ and $\sigma_{\text{out}} = 0.9$. (b), (c) The desired target patterns $\mathbf{I}_{01}$, $\mathbf{I}_{02}$.

Table 1. Wafer Stack Parameters

Layer	Index	Thickness (nm)
Incident medium	(1.45, 0)	
Top anti-reflection	(1.55, 0.0)	35
Photoresist	(1.8, 0.02)	100
Bottom anti-reflection	(1.72, 0.33)	87
Substrate	(0.833, 2.778)	

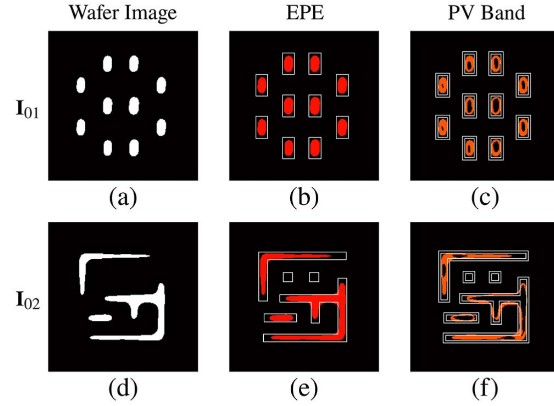

Fig. 5. Printed wafer images with (a) PE 4494 and (d) PE 5193, EPE images with (b) EPE 1158 and (e) EPE 1512, PV Band images with (c) PV Band 2347 and (f) PV Band 3965 with respect to target patterns $\mathbf{I}_{01}$ and $\mathbf{I}_{02}$ illuminated by the annular source in Fig. 4(a).

$$\boldsymbol{\phi}_{t+1} = \boldsymbol{\phi}_t - \eta_s\cdot g_t, \tag{18}$$

where $g_t = \frac{\partial\mathcal{S}}{\partial\boldsymbol{\phi}}$, the step-size η_s is set as 0.3, which is repeatedly tested for convergence, and when the proposed approach is applied, η in Eq. (15) and decay rates β_1, β_2 are suggested to be 0.1 and 0.9, 0.999.

In Fig. 6 where the proposed method and the SGD method are applied to the simulation, the columns represent the optimized source pattern $\hat{\mathbf{J}}$, the optimized mask pattern $\hat{\mathbf{M}}$, the EPE images, and the PV Band images simulated with the optimized $\hat{\mathbf{M}}$ illuminated by the optimized $\hat{\mathbf{J}}$. Two weight parameters, $\gamma_1 = \{0.6, 0.3, 0.1\}$ and $\gamma_2 = \{0.6, 0.1, 0.3\}$, are used that emphasize EPE and PV Band minimization, respectively. Figures 6(a)–6(d) show the simulation results with $\mathbf{I}_{01}$ as the target pattern, using the proposed algorithm and the SGD method weighted by γ_1 and γ_2, respectively. The values of $\mathcal{S}_{\text{pe}}$, $\mathcal{S}_{\text{epe}}$, and $\mathcal{S}_{\text{pv}}$ of the simulations in row $\mathbf{I}_{01}$ of Fig. 5 and Figs. 6(a)–6(d) are recorded in Table 2. Significant improvements of PE, EPE, and PV Band are duly observed to reduce $\mathcal{S}_{\text{pe}}$ from 4494, $\mathcal{S}_{\text{epe}}$ from 1158, and $\mathcal{S}_{\text{pv}}$ from 2347 in Fig. 5(a)–5(c) to $\mathcal{S}_{\text{pe}} = 614, 540, 586, 490$, $\mathcal{S}_{\text{epe}} = 172, 175, 174, 143$, and $\mathcal{S}_{\text{pv}} = 2246, 1834, 2211, 1885$, in Figs. 6(a)–6(d) with target pattern $\mathbf{I}_{01}$.

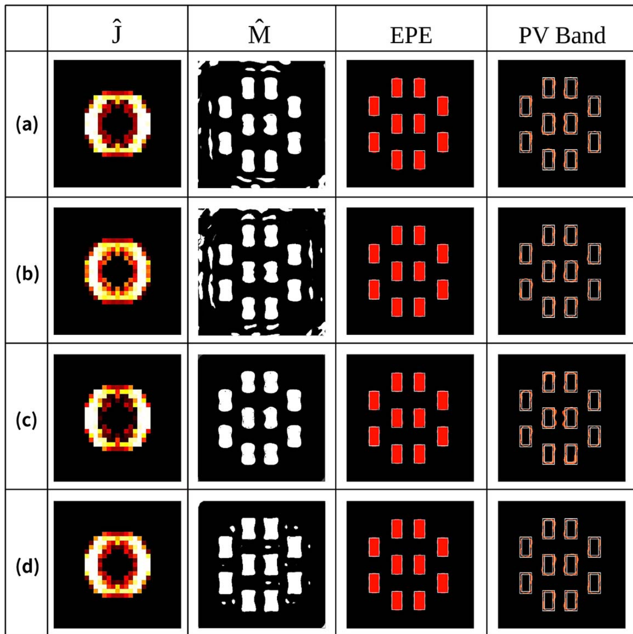

Fig. 6. Simulation results with $\mathbf{I}_{01}$ as the target pattern. Columns from left to right: the synthesized source pattern $\hat{\mathbf{J}}$, the synthesized mask pattern $\hat{\mathbf{M}}$, the EPE images, and the PV Band images illuminating $\hat{\mathbf{M}}$ by $\hat{\mathbf{J}}$. Rows: proposed approach (a) with γ_1 and (b) with γ_2, SGD (c) with γ_1 and (d) with γ_2.

Table 2. $\mathcal{S}_{\text{pe}}$, $\mathcal{S}_{\text{epe}}$, and $\mathcal{S}_{\text{pv}}$ of the Simulations in Figs. 5 and 6

	Fig. 5	Fig. 6			
	row $\mathbf{I}_{01}$	(a)	(b)	(c)	(d)
$\mathcal{S}_{\text{pe}}$	4494	614	540	586	490
$\mathcal{S}_{\text{epe}}$	1158	172	175	174	143
$\mathcal{S}_{\text{pv}}$	2347	2246	1834	2211	1885

The initial mask ω_0 in the simulations in Fig. 6 is defined as an $N \times N$ matrix with each element equaling $\pi/3$, which proves feasible for both the proposed approach and the SGD method. However, the initialization value $\omega_0 = \pi/3$ and step-size $\eta_s = 0.3$ are time-consumingly decided through many experiments, which greatly increase the workload of the simulations. Alternatively, with random initial masks ω_0 in Fig. 7, another set of simulations is performed in Fig. 8 with target pattern $\mathbf{I}_{01}$ and weight parameter γ_2 to show the impact of initial masks on the optimization process. The columns present $\hat{\mathbf{J}}$, $\hat{\mathbf{M}}$, the EPE images, and the PV Band images simulated with $\hat{\mathbf{M}}$ illuminated by $\hat{\mathbf{J}}$. Two random initial masks ω_1 and ω_2 in Figs. 7(c) and 7(d) are, respectively, applied to Figs. 8(a) and 8(b), using the proposed approach, Figs. 8(c) and 8(d) using the SGD method with weight γ_2 and target pattern $\mathbf{I}_{02}$. The values of $\mathcal{S}_{\text{pe}}$, $\mathcal{S}_{\text{epe}}$, and $\mathcal{S}_{\text{pv}}$ of the simulations in row $\mathbf{I}_{02}$ of Fig. 5 and Figs. 8(a)–8(d) are recorded in Table 3, where $n.a.$ stands for not available. It is observed

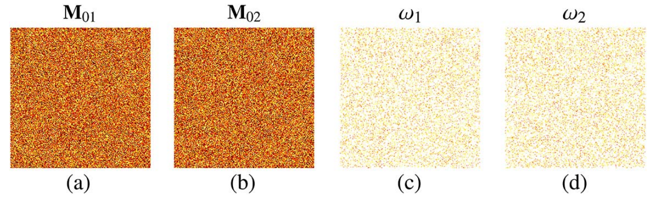

Fig. 7. Randomly initialized masks within the range $(0, 1)$; (a) $\mathbf{M}_{01}$ and (b) $\mathbf{M}_{02}$. (c) ω_1 and (d) ω_2 are the transformed parameters.

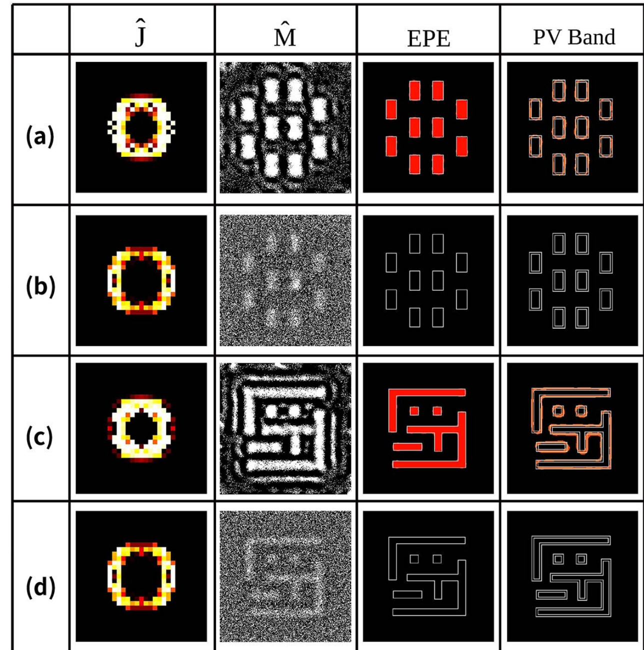

Fig. 8. Simulation results with $\mathbf{I}_{01}$ and $\mathbf{I}_{02}$ as the target pattern and weight γ_2. Rows: (a) and (c) proposed approach with ω_1 and ω_2, (b) and (d) SGD with ω_1 and ω_2 as initial masks.

Table 3. $\mathcal{S}_{\text{pe}}$, $\mathcal{S}_{\text{epe}}$, and $\mathcal{S}_{\text{pv}}$ of the Simulations in Figs. 5 and 8

	Fig. 5	Fig. 8		Fig. 5	Fig. 8	
	row $\mathbf{I}_{01}$	(a)	(b)	row $\mathbf{I}_{02}$	(c)	(d)
$\mathcal{S}_{\text{pe}}$	4494	567	$n.a.$	5193	468	$n.a.$
$\mathcal{S}_{\text{epe}}$	1158	178	$n.a.$	1512	96	$n.a.$
$\mathcal{S}_{\text{pv}}$	2347	1867	$n.a.$	3965	2472	$n.a.$

that for initial random masks ω_1 and ω_2, the proposed approach still reaches satisfactory local minimum, however, the SGD method starting with ω_1 and ω_2 finds it difficult to break symmetry to escape an unwanted local minimum resulting in poor OPC performance, showing great initial condition dependence of the SGD method.

The convergence of $\mathcal{S}$ and $\mathcal{S}_{\text{pe}}$ in the simulations in Fig. 8 is drawn in Figs. 9(a) and 9(b). In Figs. 9(c) and 9(d),

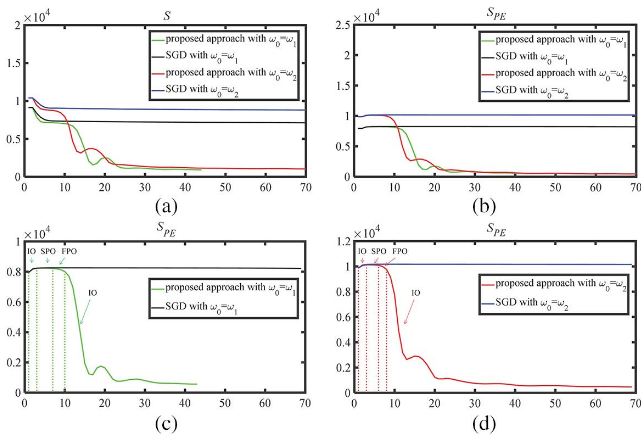

Fig. 9. Convergence of (a) $\mathcal{S}$, (b) $\mathcal{S}_{\rm pe}$ of the simulations in Fig. 8, (c) $\mathcal{S}_{\rm pe}$ of the simulations in Figs. 8(a) and 8(b), and (d) $\mathcal{S}_{\rm pe}$ of the simulations in Figs. 8(c) and 8(d).

special inspections are taken to investigate the convergence of $\mathcal{S}_{\rm pe}$ when initial masks $\mathbf{M}_{01}$ and $\mathbf{M}_{02}$ in Figs. 7(a) and 7(b) are, respectively, applied to Figs. 8(a) and 8(b), Figs. 8(c) and 8(d) with the proposed approach and the SGD method. In Figs. 9(c) and 9(d), with the SGD method, a small η_s renders very small values of $\eta_s \cdot g_t$ with random initial masks ω_1 and ω_2 and inhibits the update of ϕ_t to break symmetry when the optimization of $\mathcal{S}_{\rm pe}$ hits the local minimum, presenting very poor convergence, while a bigger η_s will lead to divergence in later iterations. On the contrary, the proposed algorithm uses bias-corrected first moment and second moment estimates $\hat{m}_t$, $\hat{v}_t$ to constrain the gradients of the objective functions, and therefore, at a certain step when the updating process reaches a local minimum, IO accumulates the moments $\hat{m}_t$, $\hat{v}_t$ and enters the FPO to attenuate $\hat{m}_t$, $\hat{v}_t$ as small enough to be close to 0 to subsequently break symmetry by entering the SPO. Such supersedure of IO, FPO, and SPO in the updating of ϕ can be observed in the Figs. 9(c) and 9(d), showing the ability of the proposed approach to escape unwanted local minima when random initial masks are applied. It should also be mentioned that the simulations in Fig. 8 present similar results for $\mathcal{S}_{\rm epe}$ and $\mathcal{S}_{\rm pv}$ with weight γ_1, showing the generality of the proposed approach.

This work was partially supported by the National Natural Science Foundation of China (No. 61875041), the Natural Science Foundation of Guangdong Province (No. 2016A030313709), and the Guangxi Science Foundation (Nos. 2013GXNSFCA019019 and 2017GXNSFAA198227).

References

1. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography* (SPIE, 2001).
2. L. W. Liebmann, S. M. Mansfield, A. K. Wong, M. A. Lavin, and T. G. Dunham, IBM J. Res. Develop. **45**, 651 (2001).
3. A. K. Wong, *Optical Imaging in Projection Lithography* (SPIE, 2005).
4. L. Pang, Y. Liu, and D. Abrams, Proc. SPIE **6823**, 68230X (2006).
5. A. Poonawala and P. Milanfar, IEEE. Trans. Image Process. **16**, 774 (2007).
6. K. Lai, A. E. Rosenbluth, S. Bagheri, J. Hoffnagle, K. Tian, D. Melville, J. Tirapu-Azpiroz, M. Fakhry, Y. Kim, S. Halle, and G. McIntyre, Proc. SPIE **7274**, 72740A (2009).
7. X. Ma and G. R. Arce, Opt. Express **17**, 5783 (2009).
8. T. Muelders, V. Domnenko, and B. Küchler, Proc. SPIE **7823**, 78233X (2010).
9. N. Jia and E. Y. Lam, in *IEEE International Conference of Electron Devices and Solid-State Circuits (EDSSC)* (2010), p. 1.
10. X. Ma, C. Han, Y. Li, L. Dong, and G. R. Arce, J. Opt. Soc. Am. A **30**, 112 (2013).
11. J. Li and E. Y. Lam, Opt. Express **22**, 9471 (2014).
12. X. Wu, S. Liu, J. Li, and E. Y. Lam, Opt. Express **22**, 3924 (2014).
13. X. Cao, M. Guan, L. Xia, X. Sang, and Z. Chen, Chin. Opt. Lett. **15**, 120901 (2017).
14. B. Zhuang, C. Xu, Y. Geng, G. Zhao, H. Chen, Z. He, Z. Wu, and L. Ren, Chin. Opt. Lett. **16**, 041102 (2018).
15. P. Zhang, D. Liu, A. Yang, and J. Zhu, Chin. Opt. Lett. **17**, 070901 (2019).
16. Z. Song, X. Ma, J. Gao, J. Wang, Y. Li, and G. R. Arce, Opt. Express **22**, 14180 (2014).
17. X. Ma, Z. Wang, Y. Li, G. R. Arce, L. Dong, and J. Garcia-Frias, Opt. Express **26**, 14479 (2018).
18. J. C. Yu and P. Yu, Proc. SPIE **7873**, 23067 (2011).
19. Y. Shen, N. Jia, and N. Wong, Opt. Express **19**, 5511 (2011).
20. Y. Shen, Opt. Express **25**, 21775 (2017).
21. Y. Shen, Opt. Express **26**, 10065 (2018).
22. N. Jia and E. Y. Lam, Opt. Express **19**, 19384 (2011).
23. J. R. Gao, X. Xu, B. Yu, and D. Z. Pan, in *IEEE Design Automation Conference* (2014), p. 1.