

Article



# Multi-Objective Defocus Robust Source and Mask Optimization Using Sensitive Penalty

# Pengzhi Wei, Yanqiu Li \*, Tie Li<sup>D</sup>, Naiyuan Sheng, Enze Li and Yiyu Sun

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; wei\_peng\_zhi@163.com (P.W.); woodlife410@163.com (T.L.); shengnaiyuan@bit.edu.cn (N.S.); lienze\_2011@126.com (E.L.); sunyiyu0105@163.com (Y.S.)

\* Correspondence: liyanqiu@bit.edu.cn

Received: 12 April 2019; Accepted: 17 May 2019; Published: 27 May 2019



**Abstract:** The continuous decrease in the size of lithographic technology nodes has led to the development of source and mask optimization (SMO) and also to the control of defocus becoming stringent in the actual lithography process. Due to multi-factor impact, defocusing is always changeable and uncertain in the real exposure process. But conventional SMO assumes the lithography system is ideal, which only compensates the optical proximity effect (OPE) in the best focus plane. Therefore, to solve the inverse lithography problem with more uniformity of pattern in different defocus variations, we proposed a defocus robust SMO (DRSMO) approach that is driven by a defocus sensitivity penalty function for the first time. This multi-objective optimization samples a wide range of defocus disturbances and it can be proceeded by the mini-batch gradient descent (MBGD) algorithm effectively. The simulation results showed that a more robust defocus source and mask can be designed through DRSMO optimization. The defocus sensitivity factor  $s_{\beta}$  maximally decreased 63.5% compared to conventional SMO, and due to the low error sensitivity and the depth of defocus (DOF), the process window (PW) was further enlarged effectively. Compared to conventional SMO, the exposure latitude (EL) maximally increased from 4.5% to 10.5% and DOF maximally increased 54.5% (EL = 5%), which proved the validity of the DRSMO method in improving the focusing performance.

**Keywords:** computational lithography; source and mask optimization (SMO); defocus robustness; process window enhancement; multi-objective optimization

## 1. Introduction

With the shrink in critical dimension (CD), the impact of the optical proximity effect (OPE) become obvious, it causes distortion in the exposure pattern and reduction of pattern fidelity and contrast, so that it must be corrected effectively. Besides, continuous shrinking also allows for control of the defocus in lithography to become increasingly stringent. In actual lithography processes, defocus is always uncertain at the wafer level, because of the unevenness of the wafer surface [1]. In addition, aberrations, thermal aberrations [2], thermal mask effects [3], and thick mask effects [4] all inevitably cause best-focus plane constant shift and further increase the OPE in off-focus conditions. Meanwhile, the continuous shrinkage of technology nodes has promoted the introduction of resolution enhancement technology (RET). Conventional RET, such as optical proximity correction (OPC) and source and mask optimization (SMO), generate the best exposure conditions by optimizing mask or simultaneously optimizing source and mask, respectively [5,6]. However, general RET methods assume the lithography system is ideal, which only compensates the OPE in the nominal condition [7–9]. But, with CD shrinkage, it results in the imaging quality becoming more sensitive to defocus,

thus requiring RET methods corresponding to off-focus conditions. Therefore, to tackle the focus variation, an analytical defocus expansion function was derived to predict the defocus aerial image in an inverse lithography technology (ILT) framework [10], so that a variational lithography model (VLIM) is derived to take into account exposure dose and focus variations [11]. Meanwhile, we previously proposed the source–mask-numerical aperture (NA) co-optimization (SMNO) method to extend the depth of defocus (DOF) by fine-tuning the NA [12,13], but it inevitably sacrificed resolution due to the reduction of NA. In addition, Peng et al. [14] also studied SMO methods to improve the pattern fidelity in the case of an assigned defocus plane which operated at 100 nm defocusing, and our subsequent works have drawn on this approach [15,16], but it is hard to ensure global fidelity in different defocus variations. Thus, Jia et al. [17–21] proposed statistical variations-based OPC and SMO methods to improve global fidelity at different defocus variations.

Unfortunately, the above methods found it hard to improve the uniformity of exposure patterns within the DOF, so that the optimized system still had relatively high sensitivity to defocusing. To eliminate changeable and uncertain defocusing due to the presence of multiple factors effectively, it requires the reduction of defocusing sensitivity for optimized systems. Therefore, to the best of our ability to minimize the defocusing sensitivity, we propose a defocus robust SMO (DRSMO) approach that is driven by a new multi-objective optimization strategy. In our method, the total cost function is composed of expectation fidelity and expectation sensitivity in different defocus disturbances. In addition, the expectation sensitivity penalty function is introduced into an inverse optimization framework for the first time to constrain the uniformity of aerial images to defocus in different disturbances, so that more robust source and mask are designed through this optimization. Compared to conventional SMO approaches, the simulation results confirmed that DRSMO can further reduce defocus sensitivity and improve process robustness. The defocus sensitivity factor  $s_{\beta}$  maximally decreased 63.5% compared to conventional SMO, and the DOF corresponding to EL = 5% maximally increased 54.5%, EL maximally increased from 4.5% to 10.5% as well. It means that larger exposure tolerances were in the actual lithography process via DRSMO. In addition, this paper also discusses the established optimization problem which was solved by the stochastic gradient descent (SGD) algorithm and mini-batch gradient descent (MBGD) algorithm. Optimization results show that due to the wide sampling range of defocus disturbances, it is easy to fall into the local optimal solution by SGD but converge well when introduced to MBGD.

The remainder of the paper is organized as follows. In Section 2, the introduction of the forward imaging model, the inverse optimization problem driven by the new multi-objective cost function, and the multi-objective optimization process by different algorithms are described. In Section 3, the simulation conditions and simulation results are presented.

## 2. DRSMO Modeling

Figure 1 illustrates the DRSMO optimization framework which is composed of the forward calculation to evaluate pattern fidelity and the inverse optimization to update the source and mask parameters. In current lithography processes, source and mask are freeform pixel-based configurations. Therefore, both the source and mask can be represented by matrix form *J* and *m*, respectively. Each source element  $J^{x_s,y_s} \in (0,1)$  represent normalized light intensity and each mask element  $m_{rs}$  is subject to 0–1 binary distribution. To overcome the complexity of the constrained optimization problem, a parameters transfer was made to convert  $J^k$  and  $m^k$  to unconstrained sources mask parameters  $\Omega_J^k$  and  $\Omega_M^k$  (see Appendix B and Equation (A9)) in the *k*th iteration, respectively [8].



Figure 1. Forward calculation and inverse optimization process for defocus robust SMO (DRSMO).

As for the forward calculation process, the printed resist pattern was calculated by the given light source and mask parameters through the corresponding physical process. Then, the printed pattern was compared with the target resist pattern to evaluate the pattern fidelity and CD error. Under the Abbe imaging principle, the aerial image that takes into account defocusing can be represented as  $I^{defocus}(\beta)$ . For the model-based SMO method, the scalar imaging model is inaccurate in hyper-NA (NA > 1) immersion lithography systems [22]. Thus, we previously studied the vector imaging model for aerial image calculation [16], and it can be formulated as

$$I^{defocus}(\beta) = \frac{1}{J_{sum}} \sum_{x_s} \sum_{y_s} \left( J^{x_s, y_s} \times \sum_{p=x, y, z} \left\| E_p^{wafer} \right\|^2 \right), \tag{1}$$

where  $J_{sum}$  is the summation of all the source intensity and is used as a normalization factor.  $E_p^{wafer}$  is the electric fields (x-, y-, and z-directions) in the exposure plane which can be expressed as

$$E_p^{wafer} = \mathcal{F}^{-1}\{2\pi/n_w \times R \times C \times V(x_s, y_s) \odot U_{ideal} \odot Def$$
  
$$\odot \mathcal{F}[M_{near}(x_s, y_s)] \odot E_i(x_s, y_s)\},$$
(2)

where  $n_w$  is the index of refraction, the magnification of R = 4 normally,  $V(x_s, y_s)$  is the vector matrix for hyper-NA systems, C is the irradiance correction factor,  $U_{ideal}$  is the ideal pupil filter,  $\mathcal{F}[M_{near}(x_s, y_s)]$  is the mask diffraction near field,  $E_i(x_s, y_s)$  is the electric field of the source that is represented by a 2 × 1 vector. Operators  $\odot$ ,  $\mathcal{F}[]$ , and  $\mathcal{F}^{-1}[]$  are represented by matrix entry-by-entry multiplication, forward Fourier transform, and inverse Fourier transform, respectively.

It should be noted that the impact of defocusing on resulting aerial image *Def* can be described as an aberration of a sort [23]. This causes a distribution of the phase change on the ideal aperture which can be described as  $\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ 

$$Def = \exp\left[j\frac{2\pi n_{w}\beta(1-y_{i})}{\lambda}\right],$$
(3)

where  $y_i$  is the direction cosine in the propagation direction and  $\beta$  is the defocus value.

Next, the resist image was adopted to the continuous derivable sig model [24], which can be expressed as

$$Z(\beta) = sig \left[ I^{defocus}(\beta) \right] = \frac{1}{1 + \exp\left\{ -a \left[ I^{defocus}(\beta) - t_r \right] \right\}}.$$
(4)

where *a* indicates the steepness of the sigmoid function, and  $t_r$  is the threshold.

#### 2.1. DRSMO Inverse Optimization Framework

As for the inverse optimization process in Figure 1, it is a continuous update source and mask parameter to meet the final target resist pattern and overcome the OPE. The inverse optimization process of DRSMO relies on the corresponding cost function establishment. In DRSMO, the multi-objective cost function is composed of the statistics expected in plentiful defocus disturbances, and the total cost function can be divided into two parts: the pattern fidelity part and defocus sensitivity part.

The pattern fidelity part is in light of Jia's [19] approach, which is defined as the expectation of the Euclidean distance between the target resistance pattern  $\tilde{Z}$  and the resistance pattern  $Z(\beta_i)$  in plentiful defocusing disturbances. It can be formulated as

$$F = \varepsilon_{\beta_i} \Big\{ \|\widetilde{Z} - Z(\beta_i)\|_2^2 \Big\},\tag{5}$$

where  $\beta_i$  is a stochastic variable representing defocusing disturbances. It is subject to a certain range of evenly distributed disturbances, namely,  $\beta_i \in U(-\alpha, \alpha)$ .  $\beta = \{\beta_i\}$  represents the whole training set. It should be noted that the sample range of  $\pm \alpha$  is selected according to the actual situation, larger  $\alpha$ can theoretically lead to wider DOFs, but too large a DOF will be beyond the potential of optimization and  $\widetilde{Z}$ ,  $Z(\beta_i) \beta_i$ , respectively.  $\varepsilon\{$  } is the mathematical expectation.

To the core ideal of DRSMO is introduced the defocus sensitivity penalty function, which aims at minimizing the expected quadratic change ratio of the aerial image to defocusing and can be formulated as

$$Y = \varepsilon_{\beta_i} \{ \|\partial I^{defocus}(\beta_i) / \partial \beta_i\|_2^2 \}, \tag{6}$$

This penalty function directly controls the change rate of the aerial image to the defocus, which improves the consistency of the pattern and CD in different defocus disturbances. Therefore, it is beneficial to optimize a more robust source and mask with a lower defocus sensitivity. Besides, due to the improvement of process robustness for the optimized system, the PW will also enlarge. Appendix A and Equation (A5) define the details about the analytical sensitivity penalty  $Y_i$ , so that the total cost function *G* consists of the weighted sum of *F* and *Y*.

$$\min G = F + \omega Y. \tag{7}$$

where  $\omega$  is the weighting factor of the sensitivity part. Typically,  $\omega = 0$  means the optimization only operates at the fidelity part *F*. In Section 3, we will discuss that the optimization results only operate at the fidelity part or simultaneously operates at the fidelity part and sensitivity part.

#### 2.2. DRSMO Optimization Algorithm

In our method, the DRSMO process can be regarded as a machine learning process, training variable  $\beta_i$  as a stochastic disturbance in the cost function to solve this multi-objective optimization problem. Table 1 illustrates the optimization flow of the SGD and MBGD algorithm, respectively. In our previous works [25,26], SGD was adopted to calculate the gradient in a single training sample  $\beta_{i,k}$  in each iteration with fast speed. However, due to the wider sample range of defocus disturbances, the SGD algorithm could not guarantee each iteration was conducted in the global optimal direction.

In order to give consideration to both optimization speed and accuracy, mini-batch gradient descent (MBGD) was proposed to traverse a part of the random defocus samples  $\beta_{i,k}$ ,  $\beta_{i+1,k}$ ,  $\cdots$ ,  $\beta_{i+l_{batch}-1,k}$  in each iteration, and  $l_{batch}$  is the batch number in each iteration [27]. Different from the SGD method, MBGD updates part of the training set, thus leading to a relatively correct search direction, so it is easier to converge to the global optimal solution.

Both the SGD and MBGD algorithm need to calculate the analytic gradient expression of cost function, which can be formulated as

$$\nabla_I G_i = \nabla_I F_i + \omega \nabla_I Y_i, \tag{8}$$

$$\nabla_M G_i = \nabla_M F_i + \omega \nabla_M Y_i. \tag{9}$$

where  $\nabla_I$  and  $\nabla_M$  are the gradient to the source parameter  $\Omega_I$  and mask pattern  $\Omega_M$ , respectively. We directed a large amount of study toward the derivation of the analytic gradient formula about sensitive penalty  $\nabla_I Y_i$ ,  $\nabla_M Y_i$ , and more details can be found in Appendix B, Equations (A11) and (A13). Similarly, the expansion of  $\nabla_I F_i$ ,  $\nabla_M F_i$  can be found in Appendix C, Equations (A19) and (A20).

Table 1. Stochastic gradient descent (SGD) and mini-batch gradient descent (MBGD) optimization procedure.

#### SGD procedure

- **1. Initialization:** Assign the starting source parameter  $\Omega_I$ , mask parameter  $\Omega_M$ , the source step size  $s_I$ , the mask step size  $s_M$ , the upper limit iteration number  $l_{smo}$
- 2. Optimization: Simultaneously update the source and mask patterns:

While  $k \leq l_{smo}$   $k \leftarrow k + 1$ Randomly generate the defocus value  $\beta_{i,k-1}$ Calculate the generate  $\nabla_I G_i^{k-1}$ ,  $\nabla_M G_i^{k-1}$ , respectively; Update the source and mask parameters

$$\Omega_{J}^{k} = \Omega_{J}^{k-1} - s_{J} \frac{\nabla_{J} G_{i}^{k-1}}{\|\nabla_{I} G_{i}^{k-1}\|_{2}}$$
$$\Omega_{M}^{k} = \Omega_{M}^{k-1} - s_{M} \frac{\nabla_{M} G_{i}^{k-1}}{\|\nabla_{M} G_{i}^{k-1}\|_{2}}$$

end

3. Output: the optimized source and mask parameters.

MBGD procedure

**1. Initialization:** Assign the starting source parameter  $\Omega_I$  mask parameter  $\Omega_M$ , the source step size  $s_I$ , the mask step size  $s_M$ , the upper limit iteration number  $l_{smo}$ , the batch number  $l_{batch}$ 

**2. Optimization:** Simultaneously update the source and mask patterns:

*While*  $k \leq l_{smo}$ 

 $k \leftarrow k + 1$ 

Random generate a set of the defocus values  $\beta_{i,k-1}, \beta_{i+1,k-1}, \dots, \beta_{i+l_{batch}-1,k-1}$ Calculate the corresponding gradient of cost function  $\nabla_J G_i^{k-1}, \nabla_J G_{i+1}^{k-1}, \dots, \nabla_J G_{i+l_{batch}-1}^{k-1}$ 

 $\nabla_M G_i^{k-1}, \nabla_M G_{i+1}^{k-1}, \cdots, \nabla_M G_{i+l_{batch}-1}^{k-1}$  respectively;

Update the source and mask parameters  $\sum_{k=0}^{k} \frac{1}{k} \sum_{i=1}^{k} \frac{\nabla_i C_i^{k-1}}{\nabla_i C_i^{k-1}}$ 

$$\Omega_{J}^{k} = \Omega_{J}^{k-1} - \frac{1}{l_{batch}} \sum_{a=i}^{l+l_{batch}-1} s_{J} \frac{|\nabla_{J}G_{a}^{k-1}||_{2}}{\|\nabla_{J}G_{a}^{k-1}\|_{2}}$$

$$\Omega_{M}^{k} = \Omega_{M}^{k-1} - \frac{1}{l_{batch}} \sum_{a=i}^{l+l_{batch}-1} s_{M} \frac{\nabla_{M}G_{a}^{k-1}}{\|\nabla_{M}G_{a}^{k-1}\|_{2}}$$
end

3. Output: the optimized source and mask parameters.

#### 3. Simulation Results and Discussion

#### 3.1. Simulation Conditions

We illustrate the DRSMO method in two test patterns as shown in Figure 2. The critical dimension (CD) of each pattern was 45 nm. Resistance patterns and mask were represented by a 201 × 201 matrix with a resolution of 5.625 nm × 5.625 nm and 22.500 nm × 22.500 nm per pixel, respectively. The imaging system parameters were set to be  $\lambda = 193$  nm and NA = 1.2. The freeform source was represented by a 21 × 21 matrix which uses TE-polarization illumination. In this paper, the whole training set  $\beta = \{\beta_i\}$  consisted of 900 random sampling points in the range of (-100 nm, 100 nm), since this sample range was extremely larger than initial DOF, which was without optimization. Taking

into account both optimization speed and accuracy, the batch number  $l_{batch}$  was set to be three per iteration and 300 iterations totally.

To evaluate the imaging fidelity, pattern error (PAE) refers to the Euclidean distance between the target pattern and the actual pattern in the resist. Generally, the smaller PAE means the higher fidelity of the lithographic imaging. It can be formulated as

$$PAE = \|\widetilde{Z} - Z(\beta)\|_{2}^{2}, \tag{10}$$

where *Z* is the binary target resist pattern and  $Z(\beta)$  is the actual resistance pattern under defocus  $\beta$ . Meanwhile, to evaluate the defocusing sensitivity quantitatively, we defined the defocus sensitivity factor  $S_{\beta}$  as the change ratio of PAE to defocusing

$$S_{\beta} = \frac{\partial PAE}{\partial \beta},\tag{11}$$

Moreover, to evaluate the process robustness in the actual exposure process, the PW was introduced to describe the restrictive relation between dose variation and focus variation. It was composed of two parameters, DOF and exposure latitude (EL). Exposure latitude is the allowable range of dose variation under a fixed defocus. Similarly, DOF is the largest acceptable defocus range under a fixed dose. Thus, PW consists of all pairs of DOF and EL which satisfy the exposure quality specification. Generally, taking the DOF when corresponding to ELs equal to 5% or 10% as process evaluation standard. Meanwhile, the PW representative calculation positions are marked at yellow lines in Figure 2.



**Figure 2.** Two test patterns used in the simulation: (**a**) Test pattern 1; (**b**) Test pattern 2. The PW calculation positions are marked at the yellow lines.

#### 3.2. Optimization Results and Analysis

In order to illustrate the negative influence of defocus, Figure 3a–f shows the optimization proceeded by initial SMO which merely operated at the best focus plane [7], and the evaluation of the printed image was under different defocus planes. Figure 3a,b show the optimized source for initial SMO and optimized mask for initial SMO, respectively. Figure 3c shows the printed image at the best focus plane, and Figure 3d–f shows the printed image under 50 nm, 70 nm, and 100 nm defocus, respectively. It clearly shows that the PAE increased extremely with an increase of defocus, proving that the initial SMO could not compensate the defocus distortion because the cost function was not involved in the defocus term. However, the defocusing error inevitably existed in the actual lithography process, thereby it was necessary to gain a better and more robust defocusing via DRSMO.



**Figure 3.** Target 2 simulation results of the initial SMO, SMO under assigned focus plane (100 nm defocusing), and DRSMO with  $\omega$  = 0.2, respectively.

Similarly, Figure 3g–l illustrates Peng's [14] SMO method which merely operates at an assigned defocus plane (100 nm defocusing). In this method, the established cost function can be formulated as the weight sum of the nominal term and defocus term

$$F = \|\widetilde{Z} - Z(0)\|_{2}^{2} + \omega \|\widetilde{Z} - Z(100)\|_{2}^{2},$$
(12)

Figure 3g,h show the optimized source and mask, respectively. Figure 3i–l shows the printed image under 0 nm, 50 nm, 70 nm, 100 nm defocus, respectively. Compared to the initial SMO, lower

distortion and PAE were acquired in each defocus plane. However, since this method merely operated at an assigned defocusing plane, the global fidelity was not so good. In Figure 3j under 50 nm and Figure 3k under 70 nm defocusing, apparent hot spots existed, shown in the center of the red circles.

Finally, the optimization results of the proposed DRSMO with  $\omega = 0.2$  are show in Figure 3m–r. It clearly shows that the distortion and PAE further declined in each defocus plane, so that more robust source and mask were designed through this optimization. Compare to SMO under an assigned defocusing plane, the DRSMO guaranteed better global fidelity in a wide range of defocus.

To further prove a robust improvement, Figure 4 depicts the defocus–PAE curves in the evaluation range of 0–100 nm for target 1 optimized systems. It should be noted that each weight factor  $\omega$ corresponds to a set of optimized source and mask. The slope of each curve reflected the process robustness, and a lower slope meant lower sensitivity for the optimized systems to focus on shifting. It should be noted that the slope of each curve gradual decreased in the order of initial SMO (blue curve), DRSMO with  $\omega = 0$  (green curve), DRSMO with  $\omega = 0.1$  (red curve), DRSMO with  $\omega = 0.2$  (azury curve), and DRSMO with  $\omega = 0.3$  (purple curve). We concluded that DRSMO is beneficial to reduce defocusing sensitivity and to gain a more uniform exposure pattern within a long range of defocusing, which means a better system robustness against uncertain and changeable focus shifts in a real lithography process. The core idea of the DRSMO is to introduce the defocusing sensitivity Y to constrain the uniformity of printed patterns in different defocus variations. Thus, simulations further compare the optimization performance of DRSMO, which merely operate at the fidelity part F ( $\omega = 0$ ) and DRSMO driven by the sensitivity penalty ( $\omega \neq 0$ ). For instance, comparing the DRSMO with  $\omega = 0$  (green curve) and DRSMO with  $\omega = 0.1$  (red curve) in Figure 4, the slope of the red curve is lower than that of the green curve, which infers that the introduction of the sensitivity penalty Y can further improve pattern uniformity with a wide range of defocus variations. It was proved that the validity of introduction sensitivity penalty Y further improves the optimization performance. In brief, to maximize the DRSMO optimization performance, both the fidelity part F and the penalty term Y must be introduced into the optimization framework, and the weight factor  $\omega$  must be chosen appropriately.



**Figure 4.** The defocus–pattern error (PAE) curves of target 1 for initial SMO (blue curve), the DRSMO with the weight factor  $\omega = 0$  (green curve),  $\omega = 0.1$  (red curve),  $\omega = 0.2$  (azury curve), and  $\omega = 0.3$  (purple curve). Optimizations were proceeded by the MBGD algorithm.

In actual lithography processes, PW is one of the critical evaluation criteria which refers to the exposure error tolerance. Figure 5 shows the PWs for the conventional SMO (blue curve), the DRSMO with the weight factor  $\omega = 0$  (green curve),  $\omega = 0.1$  (red curve),  $\omega = 0.2$  (azury curve), and  $\omega = 0.3$  (purple curve). It is illustrated that the PW of the proposed DRSMO was evidently larger than that of

the initial SMO. For the initial SMO, the maximal EL was less than 5%, which was far below the actual exposure requirements. By using DRSMO, the EL maximal increased from 4.5% to 10.5%. Similar results were found when comparing the difference of DRSMO without the sensitive penalty and DRSMO with the sensitive penalty. For example, when comparing PW with  $\omega = 0$  (green curve) and  $\omega = 0.1$  (red curve), a wider PW was found for the red curve than the green curve. It is inferred that the sensitive penalty is helpful for improving system robustness so that it indirectly boosts PW. However, because the cost function does not involve terms which directly relate to EL and DOF, the relationship between weight factor  $\omega$  and PW is uncertain and unclear. In this way, although DRSMO with  $\omega = 0.3$  had the best defocus robustness, the PW was shrunk because the weight factor was too large that it led to overfitting during the training process. This illustrates that a well-chosen weight factor  $\omega$  is important to simultaneously improve robustness and PW.



**Figure 5.** Process window (PW) of target 1 for the initial SMO (blue curve), the DRSMO with the weight factor  $\omega = 0$  (green curve),  $\omega = 0.1$  (red curve),  $\omega = 0.2$  (azury curve), and  $\omega = 0.3$  (purple curve). Optimizations were proceeded by the MBGD algorithm.

Table 2 summarizes the target 1 comparison of optimization results for conventional SMO, DRSMO with  $\omega = 0$ ,  $\omega = 0.1$ ,  $\omega = 0.2$ , and  $\omega = 0.3$ , respectively. It should be noted that PAE sensitivity factor  $S_{\beta}$  declined significantly with the  $\omega$  increase. Compare with initial SMO, the largest decrease of  $S_{\beta}$  in DRSMO with  $\omega = 0.3$  is 63.5%. Integrated consider the improvement of both  $S_{\beta}$  and PW,  $\omega = 0.1$  is a relative reasonable weight factor for maximize optimization performance.

**Table 2.** The target 1 optimized values of  $S_{\beta}$ , depth of focus (DOF) (nm) corresponding to exposure latitude (El) equal to 5% and 8% for conventional SMO, and DRSMO with  $\omega = 0, \omega = 0.1, \omega = 0.2$ , and  $\omega = 0.3$ , respectively.

| Method                   | $S_{meta}$ | DOF (EL = 5%) | DOF (EL = 8%) |
|--------------------------|------------|---------------|---------------|
| Initial SMO              | 56.1       | 0             | 0             |
| DRSMO ( $\omega = 0$ )   | 36.6       | 102           | 87            |
| DRSMO ( $\omega = 0.1$ ) | 25.5       | 122           | 107           |
| DRSMO ( $\omega = 0.2$ ) | 23.8       | 138           | 105           |
| DRSMO ( $\omega = 0.3$ ) | 20.5       | 102           | 0             |

Target 2 consisted of a series of vertical and horizontal mixed lines. Figure 6 shows the defocus–PAE curves in the evaluation range of (0 nm, 100 nm) for target 2; optimizations were proceeded by the MBGD algorithm. It should be noted that the slope of the DRSMO with a  $\omega$  = 0.2 (azury curve) lower

than that of DRSMO with a  $\omega = 0$  (green curve) and initial SMO (blue curve) proved the effectiveness of the sensitivity penalty.



**Figure 6.** The defocus–PAE curves of target 2 for initial SMO (blue curve), the DRSMO with the weight factor  $\omega = 0$  (green curve) and  $\omega = 0.2$  (azury curve), respectively. Optimizations were proceeded by the MBGD algorithm.

Meanwhile, the improvement of PW for target 2 was still apparent for the DRSMO approach. Figure 7 shows the PWs for the initial SMO (blue curve) and the DRSMO with the weight factors  $\omega = 0$  (green curve) and  $\omega = 0.2$  (azury curve). It should be noted that the PW had no significant improvement for the DRSMO with  $\omega = 0$  compared to the initial SMO, but due to the lower defocus sensitivity, the PW of the DRSMO with  $\omega = 0.2$  was enlarged. It was apparent that the introduction of the sensitivity penalty was beneficial for further improvement of the PW. Similarly, Table 3 summarizes the comparison of the target 2 optimization results for the initial SMO and DRSMO with  $\omega = 0$  and  $\omega = 0.2$ , respectively. It shows that the DOF corresponding to EL = 5% maximally increased by 54.5%.



**Figure 7.** PWs of target 2 for the initial SMO (blue curve) and the DRSMO with the weight factors  $\omega = 0$  (green curve) and  $\omega = 0.2$  (azury curve), respectively. Optimizations were proceeded by the MBGD algorithm.

| Method $S_{\beta}$       |      | DOF (EL = 5%) | DOF (EL = 8%) |  |
|--------------------------|------|---------------|---------------|--|
| Initial SMO              | 54.2 | 77            | 0             |  |
| DRSMO ( $\omega = 0$ )   | 40.3 | 82            | 0             |  |
| DRSMO ( $\omega = 0.2$ ) | 32.6 | 119           | 97            |  |

**Table 3.** The target 2 optimized values of  $S_{\beta}$ , DOF (nm) corresponding to ELs equal to 5% and 8% for the conventional SMO, and DRSMO with  $\omega = 0$ ,  $\omega = 0.1$ ,  $\omega = 0.2$ , and  $\omega = 0.3$ , respectively.

#### 3.3. Comparison of SGD and MBGD Algorithm for DRSMO

We have previously used the SGD algorithm to solve multi-objective SMO [25,26] and it converged well with fast speed. However, due to the wider sampling range of defocusing in the DRSMO framework, it was hard for the SGD algorithm to search for the global optimal direction if each iteration was only driven by one sample gradient in the training set. To compare the SGD and MBGD optimization performance for the DRSMO in terms of speed and accuracy, we generated the same training set with 900 sample points for both MBGD and SGD optimization processes (for the MBGD algorithm, there were a total of three samples per iteration and 300 iterations. For the SGD algorithm, there was a total of one sample per iteration and 900 iterations). Figure 8 illustrates that the defocus–PAE curves for target 1 optimization in the case of the same weight factor,  $\omega = 0.1$ , was proceeded by the SGD and MBGD algorithms, respectively. It was demonstrated that the slope of the DRSMO was proceeded by the MBGD (red curve), which was lower than that of the SGD (jasper curve). It indicates the better optimization performance for the MBGD in the same training set. Similarly, Figure 9 shows that the PWs with the same weight factor were proceeded by the SGD and MBGD algorithm that provided a wider PW due to the lower defocus sensitivity.



**Figure 8.** The defocus–PAE curves of target 1 for the DRSMO with the same weight factor  $\omega = 0.1$  which was proceeded by the SGD algorithm (jasper curve) and MBGD algorithms (red curve), respectively.

Table 4 summarizes the comparison of the optimization performances for target 1 proceeded by the MBGD and SGD, respectively. It clearly shows that the lower  $S_{\beta}$  and larger DOF were improved by MBGD optimization. Meanwhile, Table 4 shows the comparison of run times, although SGD had a relatively weak optimization performance but was faster in regard to convergence rates. In addition, all computations were carried out on a server with an Intel core i5 8400 CPU, 2.8GHz, 16.0 GB of RAM.



**Figure 9.** PWs of target 1 for the DRSMO with the same weight factor  $\omega = 0.1$  that were proceeded by the SGD algorithm (jasper curve) and MBGD algorithms (red curve), respectively.

In conclusion, both the MBGD and SGD were beneficial for improving DOF and PW. However, the massive samples taken from the defocusing disturbances made it hard for the SGD to converge to a global search direction. Therefore, the MBGD algorithm was applied to the DRSMO multi-objective optimization problem most effectively.

**Table 4.** The values of  $S_{\beta}$ , DOF (nm) corresponding to ELs equal to 5% and 8% and run time (seconds) for DRSMO proceeded by the MBGD and SGD, respectively.

|                          | Algorithm | $S_{\beta}$ | DOF (EL = 5%) | DOF (EL = 8%) | Run Time |
|--------------------------|-----------|-------------|---------------|---------------|----------|
| DRSMO ( $\omega = 0.1$ ) | MBGD      | 25.5        | 122           | 107           | 22,072.7 |
|                          | SGD       | 34.3        | 111           | 92            | 14,846.5 |

## 4. Conclusions

In conclusion, we proposed the DRSMO to compensate for uncertain defocus and OPE in real lithography processes. The inverse optimization framework was based on a new cost function that constrained the uniformity of an aerial image in different defocus disturbances, and a more robust lithographic source and mask with lower defocus sensitivities were designed. Using this method, the robustness against focus shifting was dramatically improved and the DOF and PW were extremely enlarged as well. It created a larger exposure tolerance in the actual lithography process and it was especially applied to high fidelity exposures in cutting-edge technical nodes.

**Author Contributions:** Conceptualization, P.W. and Y.L.; methodology, P.W. and T.L.; software, P.W. and T.L.; validation, P.W. and T.L.; formal analysis, N.S. and Y.L.; investigation, P.W. and T.L.; resources, T.L.; data curation, N.S. and Y.S.; writing—original draft preparation, P.W. and Y.L.; writing—review and editing, P.W. and Y.L.; visualization, E.L. and Y.S.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L.

**Funding:** This research was funded by the General Program of National Natural Science Foundation of China (Grant No.61675026) and the National Science and Technology Major Project (Grant No. 2017ZX02101006-001).

**Acknowledgments:** We gratefully acknowledge KLA-Tencor Corporation for providing academic use of PROLITH. We thank Mentor Graphics Corporation for providing academic use of Calibre.

Conflicts of Interest: The authors declare no conflict interest.

## Appendix A

For brevity, Equation (2),  $E_p^{wafer}$  can be simplified as:

$$E_p^{wafer} = \Upsilon_p^{x_s y_s} \otimes \Theta_{defocus},\tag{A1}$$

where

$$\Upsilon_{P}^{x_{s}y_{s}} = \frac{2\pi}{n_{w}R} \mathcal{F}^{-1}\{C \times V(x_{s}, y_{s}) \odot U_{ideal} \odot \mathcal{F}[M_{near}(x_{s}, y_{s})] \odot E_{i}(x_{s}, y_{s})\},\tag{A2}$$

$$\Theta_{defocus} = \mathcal{F}^{-1} \{ Def \},\tag{A3}$$

Similarly, according to our previous work [16],  $E_p^{wafer}$  can also be simplified as:

$$E_p^{wafer} = H_p^{x_s y_s} \otimes (B^{x_s y_s} \odot M), \tag{A4}$$

where *M* is the mask layout,  $B^{x_s y_s}$  is the mask diffraction matrix, and  $H_p^{x_s y_s}$  is the transfer function of the project lens.

Then, discretize the sensitivity penalty Y in each defocus disturbance  $Y_i$ . Based on the chain rule, according to Equations (6) and (A1), the analytical sensitivity penalty  $Y_i$  can be expressed as:

$$Y_{i} = \left( \left\| \frac{\partial I^{defocus}(\beta_{i})}{\partial \beta_{i}} \right\|_{2}^{2} \right)^{2}$$

$$= \sum_{m,n} \left( \frac{\partial I^{defocus}(\beta_{i})}{\partial \Theta^{rs}_{defocus}} \frac{\partial \Theta^{rs}_{defocus}}{\partial \beta_{i}} \right)^{2}$$

$$= \sum_{m,n} \left( \frac{2}{J_{sum}} \sum_{x_{s}} \sum_{y_{s}} J^{x_{s}, y_{s}} \sum_{rs} \left( \sum_{p=x, y, z} D_{1} \operatorname{Re}\{\Phi\} + D_{2} \operatorname{Im}\{\Phi\} \right) \right)^{2},$$
(A5)

where  $I_{mn}^{defocus}(\beta_i)$  represents the *m* and *n*th sampling point in the aerial image matrix  $I^{defocus}(\beta_i)$ , and  $\Theta_{defocus}^{rs}$  represents the *r* and *sth* sampling point in matrix  $\Theta_{defocus}$  as well. Details of  $D_1$ ,  $D_2$  and  $\Phi$  can be formulated as:

$$D_{1} = \frac{\partial \operatorname{Re}\left(\Theta_{defocus}^{rs}\right)}{\partial \beta} = -\operatorname{Re}\left\{\mathcal{F}^{-1}[\omega^{rs}\sin(\omega^{rs}\beta)]\right\} - \operatorname{Im}\left\{\mathcal{F}^{-1}[\omega^{rs}\cos(\omega^{rs}\beta)]\right\},$$
(A6)

$$D_{2} = \frac{\partial \operatorname{Im}\left(\Theta_{defocus}^{rs}\right)}{\partial \beta} = -\operatorname{Im}\left\{\mathcal{F}^{-1}[\omega^{rs}\sin(\omega^{rs}\beta)]\right\} + \operatorname{Re}\left\{\mathcal{F}^{-1}[\omega^{rs}\cos(\omega^{rs}\beta)]\right\},$$
(A7)

and

$$\Phi = \sum_{r=1}^{M} \sum_{s=1}^{N} \left( \Upsilon_{P,m-r,n-s}^{x_s y_s} \Theta_{defocus}^{rs} \right) \left( \Upsilon_{P,m-r,n-s}^{x_s y_s} \right)^*, \tag{A8}$$

# Appendix B

In order to reduce the bound-constrained source and mask optimization problem, we apply parametric transformation to realize unconstrained optimization such that [16]

$$J^{x_s, y_s} = \frac{1 + \cos \Omega_J^{x_s y_s}}{2}, m_{kl} = \frac{1 + \cos \Omega_M^{kl}}{2}.$$
 (A9)

Through this transformation, the entry values of  $\Omega_J^{x_s y_s}$ ,  $\Omega_M^{x_s y_s}$  can be enlarged in the range of  $(-\infty, \infty)$ . According to Equations (A5) and (A9), the gradient of  $Y_i$  to the source parameters  $\Omega_J^{x_s y_s}$  is

$$\frac{\partial Y_i}{\partial \Omega_J^{x_S y_S}} = \frac{\partial Y_i}{\partial J^{x_S y_S}} \frac{\partial J^{x_S y_S}}{\partial \Omega_J^{x_S y_S}} \\
= \sum_{mn} \frac{-\sin \Omega_J^{x_S y_S}}{J_{sum}} \sum_{p=x,y,z} \left( \sum_{rs} \left( \begin{array}{c} \operatorname{Re} \left\{ \left( E_p^{wafer} \right)_{mn} D_1 \left( \Upsilon_{P,m-r,n-s}^{x_S y_S} \right)^* \right\} \\
+ \operatorname{Im} \left\{ \left( E_p^{wafer} \right)_{mn} D_2 \left( \Upsilon_{P,m-r,n-s}^{x_S y_S} \right)^* \right\} \end{array} \right) \right)^2, \quad (A10)$$

Thus, simplify the matrix of Equation (A10) to form

$$\nabla_{J}Y_{i} = \frac{-\sin\Omega_{J}^{x_{s}y_{s}}}{J_{sum}} \odot \mathbf{1}_{N\times1}^{T} \left[ \sum_{P=x,y,z} \begin{pmatrix} \operatorname{Re}\left\{ \left(E_{p}^{wafer}\right) \odot \left(D_{1} \otimes \Upsilon_{P}^{x_{s}y_{s}}\right)^{*}\right\} \\ +\operatorname{Im}\left\{ \left(E_{p}^{wafer}\right) \odot \left(D_{2} \otimes \Upsilon_{P}^{x_{s}y_{s}}\right)^{*}\right\} \end{pmatrix}^{2} \right] \mathbf{1}_{N\times1},$$
(A11)

where  $1_{N \times 1}$  is the one-valued vector.

Similarly, according to Equations (A5) and (A9), the gradient of  $Y_i$  to the mask parameters  $\Omega_M^{x_s y_s}$  can be formulated as

$$\frac{\partial Y_{i}}{\partial \Omega_{M}^{kl}} = \frac{\partial Y_{i}}{\partial m_{kl}} \frac{\partial m_{kl}}{\partial \Omega_{M}^{kl}} \\
= \sum_{mn} \frac{-2\sin\Omega_{M}^{kl}}{J_{sum}} \times \sum_{x_{s}} \sum_{y_{s}} J^{x_{s},y_{s}} \sum_{p=x,y,z} \{X_{mn} \\
\left[ \operatorname{Re}\left\{ \left( D_{1} \otimes \Upsilon_{p}^{x_{s}y_{s}} \right)^{*} h_{p,m-k,n-l}^{x_{s}y_{s}} - h_{kl}^{x_{s}y_{s}} + \left( h_{1p,m-k,n-l}^{x_{s}y_{s}} b_{kl}^{x_{s}y_{s}} \right)^{*} \left( \Upsilon_{p}^{x_{s}y_{s}} \otimes \Theta_{defocus} \right) \right\} + \\
\operatorname{Im}\left\{ \left( D_{2} \otimes \Upsilon_{p}^{x_{s}y_{s}} \right)^{*} h_{p,m-k,n-l}^{x_{s}y_{s}} - h_{kl}^{x_{s}y_{s}} + \left( h_{2p,m-k,n-l}^{x_{s}y_{s}} b_{kl}^{x_{s}y_{s}} \right)^{*} \left( \Upsilon_{p}^{x_{s}y_{s}} \otimes \Theta_{defocus} \right) \right\} \right] \right\},$$
(A12)

Thus, simplify the matrix to form

$$\begin{split} \nabla_{M}Y_{i} &= \frac{-2\sin\Omega_{M}}{l_{sum}} \odot \sum_{x_{s}} \sum_{y_{s}} J^{x_{s},y_{s}} \sum_{p=x,y,z} \left( \operatorname{Re}\left\{ \left[ \left( X \odot \left( D_{1} \otimes \Upsilon_{p}^{x_{s}y_{s}} \right)^{*} \right) \otimes h_{p}^{x_{s},y_{s}^{\circ}} \right] \odot B^{x_{s},y_{s}} \right. \right. \\ & \left. + \left[ \left( X \odot E_{p}^{wafer} \right) \otimes h_{1p}^{x_{s},y_{s}^{*\circ}} \right] \odot B^{x_{s},y_{s}^{*}} \right\} + \operatorname{Im}\left\{ \left[ \left( X \odot \left( D_{2} \otimes \Upsilon_{p}^{x_{s}y_{s}} \right)^{*} \right) \otimes h_{p}^{x_{s},y_{s}^{\circ}} \right] \odot B^{x_{s},y_{s}} \right. \right. \end{split}$$
(A13)  
$$& \left. + \left[ \left( X \odot E_{p}^{wafer} \right) \otimes h_{2p}^{x_{s},y_{s}^{*\circ}} \right] \odot B^{x_{s},y_{s}^{*}} \right\} \right), \end{split}$$

where *o* rotates the matrix in the argument by 180°, and \* is the conjugate operator, respectively. For brevity in the description, we simplified the above terms  $h_p^{x_sy_s}$ ,  $h_{1p}^{x_sy_s}$ ,  $h_{2p}^{x_sy_s}$ , X, and can be expanded

$$h_p^{x_s y_s} = H_p^{x_s y_s} \otimes \Theta_{defocus}, \tag{A14}$$

$$h_{1p}^{x_s y_s} = H_p^{x_s y_s} \otimes D_1,$$
 (A15)

$$h_{2p}^{x_s y_s} = H_p^{x_s y_s} \otimes D_2,$$
 (A16)

$$X = \sum_{rs} D_1^{rs} \operatorname{Re}\{\Phi\} + D_2^{rs} \operatorname{Im}\{\Phi\},$$
(A17)

## Appendix C

Discretize the fidelity part *F* in each defocus disturbance. Actually, each gradient of fidelity part  $F_i$  has already been derived in our previous work [16], the gradient of  $F_i$  to the source parameters  $\Omega_J^{x_s y_s}$  is

$$\frac{\partial F_i}{\partial \Omega_J^{x_s y_s}} = \frac{\partial F_i}{\partial Z_{mn}} \frac{\partial Z_{mn}}{\partial J^{x_s, y_s}} \frac{\partial I^{x_s, y_s}}{\partial \Omega_J^{x_s y_s}} \\ = \frac{a \sin \Omega_J^{x_s y_s}}{J_{sum}} \sum_{n=1}^N \sum_{m=1}^M (\widetilde{Z}_{mn} - Z_{mn}) (1 - Z_{mn}) Z_{mn} \times \sum_{p=x, y, z} \left| \sum_{r=1}^N \sum_{s=1}^M \Upsilon_{P, m-r, n-s}^{x_s y_s} \Theta_{defocus}^{rs} \right|^2,$$
(A18)

Thus, simplify the matrix to form

$$\nabla_{J}F_{i} = \frac{a \sin \Omega_{J}}{J_{sum}} \odot \mathbf{1}_{N\times 1}^{T} \left[ \sum_{p=x,y,z} \left\| E_{p}^{wafer} \right\|^{2} \odot \left( \widetilde{Z} - Z \right) F_{i} \right] \\ \odot Z \odot (1-Z) \mathbf{1}_{N\times 1},$$
(A19)

Likewise, the gradient of  $F_i$  to the mask parameters  $\Omega_M^{x_s y_s}$  is

$$\nabla_M F_i = \frac{2a \sin \Omega_M}{J_{sum}} \odot \sum_{x_s} \sum_{y_s} \left[ J^{x_s, y_s} \sum_{p=x, y, z} \operatorname{Re} \left\{ B^{x_s y_s *} \odot \left[ h_p^{x_s y_s *} \otimes \Lambda_p \right] \right\} \right], \tag{A20}$$

where  $\Lambda_p = E_p^{wafer} \odot \left( \widetilde{Z} - Z \right) \odot Z \odot (1 - Z).$ 

## References

- Fujisawa, T.; Asano, M.; Sutani, T.; Inoue, S.; Yamada, H.; Sugamoto, J.; Okumura, K.; Hagiwara, T.; Oka, S. Wafer flatness for CD control in photolithography. In Proceedings of the SPIE's 27th Annual International Symposium on Microlithography, Santa Clara, CA, USA, 3–8 March 2002; Volume 4691.
- Mao, Y.; Li, S.; Sun, G.; Wang, J.; Duan, L.; Bu, Y.; Wang, X. The thermal aberration analysis of a lithography projection lens. In Proceedings of the SPIE Advanced Lithography, San Jose, CA, USA, 26 February–2 March 2017; Volume 10147.
- Khounsary, A.M.; Chojnowski, D.; Mancini, D.C.; Lai, B.P.; Dejus, R.J. Thermal management of masks for deep x-ray lithography. In Proceedings of the Optical Science, Engineering and Instrumentation '97, San Diego, CA, USA, 27 July–1 August 1997; Volume 3151.
- 4. Azpiroz, J.T.; Rosenbluth, A.E. Impact of Sub-Wavelength Electromagnetic Diffraction in Optical Lithography for Semiconductor Chip Manufacturing; IEEE: New York, NY, USA, 2013.
- Rosenbluth, A.E.; Bukofsky, S.; Hibbs, M.; Lai, K.F.; Molless, A.; Singh, R.N.; Wong, A. Optimum mask and source patterns to print a given shape. In *Optical Microlithography XIV*, *Pts 1 and 2*; Progler, C.J., Ed.; Spie-Int Soc Optical Engineering: Bellingham, WA, USA, 2001; Volume 4346, pp. 486–502.
- Ma, X.; Arce, G.R. Pixel-based OPC optimization based on conjugate gradients. *Opt. Express* 2011, 19, 2165–2180. [CrossRef] [PubMed]
- 7. Ma, X.; Arce, G.R. Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography. *Opt. Express* **2009**, *17*, 5783–5793. [CrossRef] [PubMed]
- 8. Ma, X.; Han, C.Y.; Li, Y.Q.; Dong, L.S.; Arce, G.R. Pixelated source and mask optimization for immersion lithography. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **2013**, *30*, 112–123. [CrossRef] [PubMed]
- 9. Yu, J.C.; Yu, P.C. Gradient-Based Fast Source Mask Optimization (SMO). In *Optical Microlithography XXIV*; Dusa, M.V., Ed.; Spie-Int Soc Optical Engineering: Bellingham, WA, USA, 2011; Volume 7973.
- Yu, P.; Pan, D.Z.; Mack, C.A. Fast lithography simulation under focus variations for OPC and layout optimizations. In *Design and Process Integration for Microelectronic Manufacturing IV*; Wong, A.K.K., Singh, V.K., Eds.; Spie-Int Soc Optical Engineering: Bellingham, WA, USA, 2006; Volume 6156.
- Yu, P.; Shi, S.X.; Pan, D.Z. True process variation aware optical proximity correction with variational lithography modeling and model calibration. *J. Micro-Nanolithogr. MEMS MOEMS* 2007, *6*, 574–576. [CrossRef]
- 12. Guo, X.J.; Li, Y.Q.; Dong, L.S.; Liu, L.H.; Ma, X.; Han, C.Y. Parametric source-mask-numerical aperture co-optimization for immersion lithography. *J. Micro-Nanolithogr. MEMS MOEMS* **2014**, *13*, 043013. [CrossRef]
- 13. Sheng, N.; Li, E.; Sun, Y.; Li, T.; Li, Y.; Wei, P.; Liu, L. Mitigating the Impact of Mask Absorber Error on Lithographic Performance by Lithography System Holistic Optimization. *Appl. Sci.* **2019**, *9*, 1275. [CrossRef]
- 14. Peng, Y.; Zhang, J.Y.; Wang, Y.; Yu, Z.P. Gradient-Based Source and Mask Optimization in Optical Lithography. *IEEE Trans. Image Process.* **2011**, *20*, 2856–2864. [CrossRef] [PubMed]
- Ma, X.; Li, Y.Q.; Guo, X.J.; Dong, L.S. Robust Resolution Enhancement Optimization Methods to Process Variations based on Vector Imaging Model. In *Optical Microlithography XXV, Pts 1and 2*; Conley, W., Ed.; Spie-Int Soc Optical Engineering: Bellingham, WA, USA, 2012; Volume 8326.

- 16. Ma, X.; Li, Y.Q.; Guo, X.J.; Dong, L.S.; Arce, G.R. Vectorial mask optimization methods for robust optical lithography. *J. Micro-Nanolithogr. MEMS MOEMS* **2012**, *11*, 043008. [CrossRef]
- 17. Jia, N.; Lam, E.Y. Machine learning for inverse lithography: Using stochastic gradient descent for robust photomask synthesis. *J. Opt.* **2010**, *12*. [CrossRef]
- 18. Jia, N.N.; Lam, E.Y. Pixelated source mask optimization for process robustness in optical lithography. *Opt. Express* **2011**, *19*, 19384–19398. [CrossRef] [PubMed]
- Jia, N.N.; Wong, A.K.; Lam, E.Y. Robust Mask Design with Defocus Variation Using Inverse Synthesis. In *Lithography Asia 2008*; Chen, A.C., Lin, B., Yen, A., Eds.; Spie-Int Soc Optical Engineering: Bellingham, WA, USA, 2008; Volume 7140.
- 20. Li, S.K.; Wang, X.Z.; Bu, Y. Robust pixel-based source and mask optimization for inverse lithography. *Opt. Laser Technol.* **2013**, *45*, 285–293. [CrossRef]
- 21. Shen, Y.J.; Jia, N.N.; Wong, N.; Lam, E.Y. Robust level-set-based inverse lithography. *Opt. Express* **2011**, *19*, 5511–5521. [CrossRef] [PubMed]
- 22. Gallatin, G.M. High-numerical-aperture scalar imaging. Appl. Opt. 2001, 40, 4958–4964. [CrossRef] [PubMed]
- 23. Mack, C. Fundamental Principles of Optical Lithography: The Science of Microfabrication; Wiley: Hoboken, NJ, USA, 2007; pp. 265–276.
- 24. Ma, X.; Arce, G. Binary mask optimization for inverse lithography with partially coherent illumination. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* **2008**, *25*, 2960–2970. [CrossRef] [PubMed]
- 25. Han, C.Y.; Li, Y.Q.; Ma, X.; Liu, L.H. Robust hybrid source and mask optimization to lithography source blur and flare. *Appl. Opt.* **2015**, *54*, 5291–5302. [CrossRef] [PubMed]
- 26. Li, T.; Li, Y.Q. Lithographic Source and Mask Optimization with Low Aberration Sensitivity. *IEEE Trans. Nanotechnol.* **2017**, *16*, 1099–1105. [CrossRef]
- Li, M.; Zhang, T.; Chen, Y.; Smola, A.J. Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '14, New York, NY, USA, 24–27 August 2014; pp. 661–670.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).