

Pixelated source mask optimization for process robustness in optical lithography

Ningning Jia and Edmund Y. Lam*

*Imaging Systems Laboratory, Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong*

*elam@eee.hku.hk

Abstract: Optical lithography has enabled the printing of progressively smaller circuit patterns over the years. However, as the feature size shrinks, the lithographic process variation becomes more pronounced. Source-mask optimization (SMO) is a current technology allowing a co-design of the source and the mask for higher resolution imaging. In this paper, we develop a pixelated SMO using inverse imaging, and incorporate the statistical variations explicitly in an optimization framework. Simulation results demonstrate its efficacy in process robustness enhancement.

© 2011 Optical Society of America

OCIS codes: (110.3960) Microlithography; (110.5220) Photolithography; (110.1758) Computational imaging.

References and links

1. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography* (SPIE, 2001).
2. M. Rothschild, "A roadmap for optical lithography," *Opt. Photon. News* **21**(6), 26–31 (2010).
3. E. Y. Lam and A. K. Wong, "Computation lithography: virtual reality and virtual virtuality," *Opt. Express* **17**(15), 12259–12268 (2009).
4. A. Poonawala and P. Milanfar, "Mask design for optical microlithography — an inverse imaging problem," *IEEE Trans. Image Process.* **16**(3), 774–788 (2007).
5. L. Pang, Y. Liu, and D. Abrams, "Inverse lithography technology (ILT): a natural solution for model-based SRAF at 45nm and 32nm;" in *Photomask and Next-Generation Lithography Mask Technology XIV*, vol. 6607 of *Proc. SPIE*, p. 660739 (2007).
6. Y. Shen, N. Wong, and E. Y. Lam, "Level-set-based inverse lithography for photomask synthesis," *Opt. Express* **17**(26), 23690–23701 (2009).
7. X. Ma and G. R. Arce, "Generalized inverse lithography methods for phase-shifting mask design," *Opt. Express* **15**(23), 15066–15079 (2007).
8. S. H. Chan, A. K. Wong, and E. Y. Lam, "Initialization for robust inverse synthesis of phase-shifting masks in optical projection lithography," *Opt. Express* **16**(19), 14,46–14760 (2008).
9. N. Jia and E. Y. Lam, "Machine learning for inverse lithography: using stochastic gradient descent for robust photomask synthesis," *J. Opt.* **12**(4), 045601 (2010).
10. Y. Shen, N. Jia, N. Wong, and E. Y. Lam, "Robust level-set-based inverse lithography," *Opt. Express* **19**(6), 5511–5521 (2011).
11. M. Burkhardt, A. Yen, C. Proglar, and G. Wells, "Illuminator design for the printing of regular contact patterns," *Microelectron. Eng.* **41–42**, 91–96 (1998).
12. R. Socha, M. Eurlings, F. Nowak, and J. Finders, "Illumination optimization of periodic patterns for maximum process window," *Microelectron. Eng.* **61–62**, 57–64 (2002).
13. A. E. Rosenbluth, S. Bukofsky, C. Fonseca, M. Hibbs, K. Lai, R. N. Singh, and A. K. Wong, "Optimum mask and source patterns to print a given shape," *J. Microlith. Microfab. Microsys.* **1**(1), 13–30 (2002).
14. T. Fühner, A. Erdmann, and S. Seifert, "Direct optimization approach for lithographic process conditions," *J. Microlith. Microfab. Microsys.* **6**(3), 031006 (2007).
15. K. Lai, S. Bagheri, K. Tian, J. Tirapu-Azpiroz, S. Halle, G. McIntyre, D. Corliss, A. E. Rosenbluth, D. Melville, A. Wagner, M. Burkhardt, J. Hoffnagle, Y. Kim, G. Burr, M. Fakhry E. Gallagher, T. Faure, M. Hibbs, D. Flagello, J. Zimmermann, B. Kneer, F. Rohmund, F. Hartung, C. Hennerkes, M. Maul, R. Kazinczi, A. Engelen, R. Carpaaj, R. Groenendijk, J. Hageman, and C. Russ, "Experimental result and simulation analysis for the use of pixelated

- illumination from source mask optimization for 22nm logic lithography process,” in *Optical Microlithography XXII*, vol. 7274 of *Proc. SPIE*, p. 72740A (2009).
16. L. Pang, P. Hu, D. Peng, D. Chen, T. Cecil, L. He, G. Xiao, V. Tolani, T. Dam, K.-H. Baik, and B. Gleason, “Source mask optimization (SMO) at full chip scale using inverse lithography technology (ILT) based on level set methods,” in *Lithography Asia 2009*, vol. 7520 of *Proc. SPIE*, p. 75200X (2009).
 17. T. Mülders, V. Domnenko, B. Küchler, T. Klimpel, H.-J. Stock, A. Poonawala, K. N. Taravade, and W. A. Stanton, “Simultaneous source-mask optimization: a numerical combining method,” in *Photomask Technology 2010*, vol. 7823 of *Proc. SPIE*, p. 78233X (2010).
 18. X. Ma and G. R. Arce, “Pixel-based simultaneous source and mask optimization for resolution enhancement in optical lithography,” *Opt. Express* **17**(7), 5783–5793 (2009).
 19. J.-C. Yu and P. Yu, “Gradient-based fast source mask optimization (SMO),” in *Optical Microlithography XXIV*, vol. 7973 of *Proc. SPIE*, p. 797320 (2011).
 20. Y. Peng, J. Zhang, Y. Wang, and Z. Yu, “Gradient-based source and mask optimization in optical lithography,” *IEEE Trans. Image Process.* **99**, 1–10 (2011).
 21. A. K. Wong, *Optical Imaging in Projection Microlithography* (SPIE, 2005).
 22. Y. Granik, “Source optimization for image fidelity and throughput,” *J. Microlith. Microfab. Microsys.* **3**(4), 509–522 (2004).
 23. J.-C. Yu and P. Yu, “Impacts of cost functions on inverse lithography patterning,” *Opt. Express* **18**(22), 23331–23342 (2010).
 24. D. Strong and T. Chan, “Edge-preserving and scale-dependent properties of total variation regularization,” *Inverse Probl.* **19**(6), 165–187 (2003).
 25. M. K. Ng, H. Shen, E. Y. Lam, and L. Zhang, “A total variation regularization based super-resolution reconstruction algorithm for digital video,” *EURASIP Journal on Advances in Signal Processing* **2007**, Article ID 74,585 (2007).
 26. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. (Prentice Hall, 2002).
 27. C. Mack, *Fundamental Principles of Optical Lithography: The Science of Microfabrication* (Wiley, 2007).
 28. N. Jia, A. K. Wong, and E. Y. Lam, “Robust mask design with defocus variation using inverse synthesis,” in *Lithography Asia*, vol. 7140 of *Proc. SPIE*, p. 71401W (2008).
 29. J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, 2006).
 30. N. Jia and E. Y. Lam, “Performance analysis of pixelated source-mask optimization for optical microlithography,” in *IEEE International Conference on Electron Devices and Solid-State Circuits* (2010).
 31. L. Pang, G. Xiao, V. Tolani, P. Hu, T. Cecil, T. Dam, K.-H. Baik, and B. Gleason, “Considering MEEF in inverse lithography technology (ILT) and source mask optimization (SMO),” in *Photomask Technology*, H. Kawahira and L. S. Zurbrick, eds., vol. 7122 of *Proc. SPIE*, p. 71221W (2008).
 32. R. J. Socha, D. J. Van Den Broeke, S. D. Hsu, J. F. Chen, T. L. Laidig, N. P. Corcoran, U. Hollerbach, K. E. Wampler, X. Shi, and W. E. Conley, “Contact hole reticle optimization by using interference mapping lithography (IML),” in *Photomask and Next-Generation Lithography Mask Technology XI*, H. Tanabe, ed., vol. 5446 of *Proc. SPIE*, pp. 516–534 (2004).

1. Introduction

Optical lithography has served the semiconductor industry for decades as the predominant microlithography technology. This is attributed to the continuous technology development for shorter exposure wavelength and larger numerical aperture (NA) to achieve smaller minimum printed feature size [1]. In addition, resolution enhancement techniques (RETs) are developed and become essential for maintaining good printed image quality. Nowadays, as the optical lithography has entered the low- k_1 regime [2], printed feature dimensions are highly sensitive to process variations. Traditional RETs are inadequate for the dual task of printing small features and providing enough process margins, which triggers the emergence of more aggressive techniques with new computational strategies. In particular, optimization and image processing techniques participate in more advanced optical proximity correction (OPC) and illumination modification approaches to enrich the lithographers’ arsenal [3].

The objective of OPC is to compensate the undesired distortions on printed images by deliberately introducing pre-distortions on the mask geometrical shapes. An approach under active research is inverse lithography technology (ILT), which promises to deliver superior performance by enlarging the search space for the mask pattern, using computational techniques such as gradient-based mask optimization [4] and the level-set method [5, 6]. The methods can work

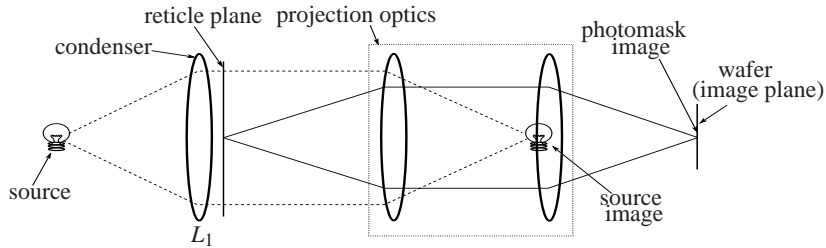


Fig. 1. Illustration of an optical projection lithography system [1, 3].

on binary and phase-shifting masks (PSMs) [7, 8], and can increase the robustness of the resulting design [9, 10]. Meanwhile, another RET known as illumination modification collects more diffraction orders by adjusting the illumination shapes [1], resulting in designs beyond traditional circular or annular shapes. Early work includes determining the optimal configuration using the diffraction orders of regular contact arrays [11], and choosing important source areas for image contrast enhancement using Hopkins partially coherent imaging equations [12].

Recently, source design and the reticle pattern optimization have been integrated and optimized together. Rosenbluth et al. [13] decomposed the source by arcs, and developed a set of constraints to compute the optimum source and mask with the maximum exposure latitude. Fühner et al. [14] adopted a more flexible meshpoint illumination representation defined by track/sector in their genetic optimization framework with the consideration of different process conditions. Currently, customized diffractive optical element (DOE) realizes a pixelated source, where the intensity and shape can be freely adjusted, therefore providing more degrees of freedom for optimization [15]. Together with the pixelated mask, the so-called free-form source-mask optimization (SMO) fits well into the inverse lithography framework [16, 17]. A gradient-based SMO algorithm was shown to improve the pattern fidelity at the specified imaging conditions [18], while another scheme implicitly considered dose sensitivity [19], yet the main focus is still the pattern fidelity at the best process conditions, and their results are obtained from two separate optimization steps: source optimization and the successive mask optimization. In the SMO framework proposed by Peng et al. [20], process robustness is considered by incorporating only one defocus condition rather than the dose-focus matrix.

In this paper, we design robust free-form source and mask patterns with respect to process variations using inverse imaging. To achieve this, we develop a cost function that incorporates not only the pattern fidelity but also the aerial image intensity distribution. We then build a statistical SMO framework and solve it by alternating optimizations of the source and the mask.

2. Lithography imaging model

An optical lithography imaging system is depicted in Fig. 1. The reticle, or the photomask, is illuminated by a light source through a condenser lens L_1 . The projection optics then forms an image of the photomask onto the wafer. Due to diffraction and different optical aberrations, however, this is necessarily a distorted image; the goal of inverse lithography is to design, through mathematical modeling and computations, the mask pattern — and sometimes the source as well — so as to achieve a desired printed image.

Detailed analysis of the lithography imaging system model has been developed over the years. Here, our focus is to introduce to the readers how the image on the wafer is distorted from the mask pattern. Let the former be $I(x, y)$ and the latter be $M(x, y)$. It is also useful to define the frequency domain representation of the mask; hence, we denote the mask pattern spectrum by $\hat{M}(f, g)$, where f and g are normalized frequency variables [21, p. 67]. Two other quantities

are also important to describe the lithography system. The first is the optical transfer function denoted by $\hat{H}(f, g)$, where its inverse Fourier transform, called the point spread function, is $H(x, y)$. The second is the effective light source $\hat{J}(f, g)$ (the Fourier transform of the mutual intensity), which arises because lithography systems involve partially coherent imaging. With these quantities, the intensity distribution at the wafer can be described by [21, Eq. 4.35]

$$I_a(x, y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{J}(f, g) \hat{H}(f + f', g + g') \hat{H}^\dagger(f + f'', g + g'') \hat{M}(f', g') \hat{M}^\dagger(f'', g'') \times e^{-i2\pi[(f' - f'')x + (g' - g'')y]} df dg df' dg' df'' dg'', \quad (1)$$

where \dagger denotes complex conjugate. The six-fold integration can be simplified to

$$I_a(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{J}(f, g) \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{H}(f + f', g + g') \hat{M}(f', g') e^{-i2\pi(f'x + g'y)} df' dg' \right|^2 df dg \approx \sum_{f, g} \hat{J}(f, g) |M(x, y) * H(x, y; f, g)|^2, \quad (2)$$

where the approximation is needed for computation in the discrete domain.

The above accounts for the light intensity arriving at the image plane, also known as the aerial image, but this is not what is printed on the wafer. Light reacts with the photoresist, which either increases its development rate with exposure (positive resists) or decreases its rate (negative resists). An image is formed at a particular location when the development is beyond a certain threshold. Thus, $I(x, y)$ is the binarized version of $I_a(x, y)$. For numerical considerations, however, we often avoid using a hard threshold in computing $I(x, y)$. Instead, a smooth transition is preferred. A frequently used model is with the sigmoid function, given by

$$I(x, y) = \text{sig}\{I_a(x, y)\} = \frac{1}{1 + e^{-\alpha(I_a(x, y) - t)}}, \quad (3)$$

where t is the threshold and α controls the steepness of the transition. Combining Eq. (2) and (3), the lithography imaging model is therefore

$$I(x, y) = \text{sig}\left\{ \sum_{f, g} \hat{J}(f, g) |M(x, y) * H(x, y; f, g)|^2 \right\}. \quad (4)$$

3. Source mask optimization framework

With a fixed optical setup, we can observe from Eq. (4) that the resulting image on the wafer is controlled by two variables: the mask pattern $M(x, y)$, and the source, which governs the mutual intensity function $\hat{J}(f, g)$. In principle, to determine if a pattern can be printed at all, and if so, what the proper light source and mask pattern should be, we should investigate all combinations to see if we can arrive at the desired $\bar{I}(x, y)$. This is the rationale behind SMO.

Unfortunately, we also observe from Eq. (4) that the image is nonlinear in $M(x, y)$ and $\hat{J}(f, g)$. Practically, we allow errors in the resulting printed image from our desired pattern; we are content if they do not cause intolerable changes in the resulting circuit's behavior. Let us denote the desired pattern with $\bar{I}(x, y)$. Furthermore, assume that we are designing a binary mask, so $M(x, y)$ can either be zero or one at any location. We can let it take on other values if we are interested in designing PSMs. Thus, we solve the following optimization problem

$$\begin{aligned} & \text{minimize} && \mathcal{D}\{I(x, y), \bar{I}(x, y)\} \\ & \text{subject to} && M(x, y) \in \{0, 1\} \\ & && \hat{J}(f, g) \geq 0. \end{aligned} \quad (5)$$

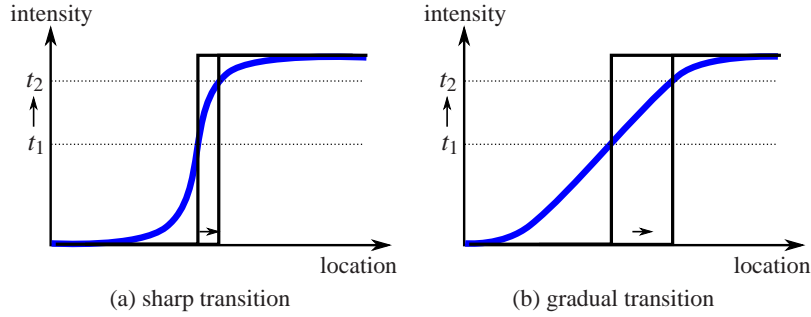


Fig. 2. Relationship among intensity distributions, threshold, and the binarized image.

The operator $\mathcal{D}\{a,b\}$ measures dissimilarity between a and b . Various formulas have been proposed for different specifications [22, 23]. Here, we define it to consist of four terms, i.e.,

$$\mathcal{D}\{I(x,y),\bar{I}(x,y)\} = \mathcal{T}\{I(x,y),\bar{I}(x,y)\} + \gamma_1 \mathcal{R}_{\text{TV}}\{I_a(x,y)\} + \gamma_2 \mathcal{R}_{\text{aerial}}\{I_a(x,y),\bar{I}(x,y)\} + \gamma_3 \mathcal{R}_{\text{contrast}}\{I_a(x,y),\bar{I}(x,y)\}. \quad (6)$$

The first term, $\mathcal{T}\{a,b\}$, ensures pattern fidelity, while the rest are regularization terms controlled by γ_1 , γ_2 and γ_3 respectively. We explain each of them below.

3.1. Pattern fidelity

The pattern fidelity term $\mathcal{T}\{a,b\}$ is used to count the errors, or mismatches, between a and b , summing over all locations. By putting this as a penalty in the optimization, the printed contour deviation from the desired one is minimized to achieve the smallest accumulated edge placement error (EPE) over the image.

For mathematical convenience, the ℓ_1 and the square of the ℓ_2 norms are frequently used, i.e.,

$$(\ell_1 \text{ norm}) \quad \mathcal{T}\{I(x,y),\bar{I}(x,y)\} = \sum_{x,y} |I(x,y) - \bar{I}(x,y)| \quad (7)$$

$$(\ell_2 \text{ norm}) \quad \mathcal{T}\{I(x,y),\bar{I}(x,y)\} = \sum_{x,y} [I(x,y) - \bar{I}(x,y)]^2. \quad (8)$$

In principle, we can use a weighted norm, where the weight is proportional to the extent that an error is allowed at the location. For example, we want to severely penalize any location that may result in bridging two disjoint areas, but an error in an isolated region may often be acceptable. However, this requires further image understanding and analysis and possibly some understanding of the underlying circuit design, and is generally difficult to accomplish. For the experiments described in this manuscript, we use the ℓ_2 norm.

3.2. Smoothing

While the pattern fidelity criterion above is concerned with the binarized image printed on the wafer only, we need to take into account the aerial image, $I_a(x,y)$, in the optimization process as well. One important reason is due to process variations. Consider Fig. 2, which plots two possible intensity distributions as a function of the spatial locations. With the same threshold (t_1), the resulting binarized images are identical. However, suppose there exists variations in the exposure, and consequently the threshold is now at t_2 . This causes little change in the printed image for (a), where the transition is sharp, but a significant deviation for (b), which the transition is gradual. Thus, it is desirable to have sharp transitions to make the design robust.

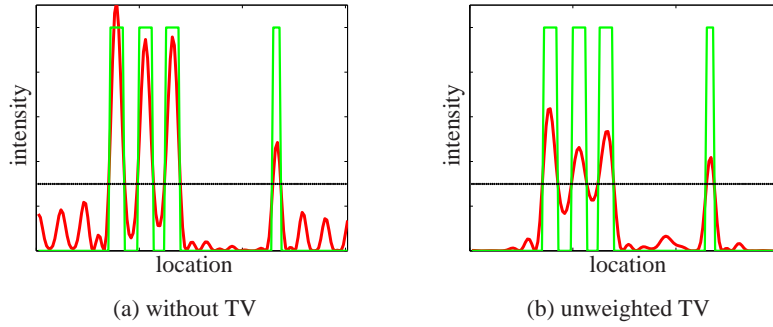


Fig. 3. Demonstration of TV regularization. The red and green curves represent aerial image intensity and target design, respectively. The dotted line marks the threshold t .

To quantify this, we argue that $I_a(x, y)$ should be piecewise smooth with sharp edges at the transition regions, and therefore the mathematical operation called total variation (TV) is the appropriate metric to use. It is now a common tool in image reconstruction and restoration, and is known to suppress the small-scale noise while preserving the large-scale features [24, 25]. The TV norm is given by

$$\|I_a(x, y)\|_{\text{TV}} = \sum_{x, y} |\nabla I_a(x, y)|, \quad (9)$$

where $\nabla I_a(x, y)$ denotes the gradient of $I_a(x, y)$. In numerical implementations, this gradient is approximated by finite difference. Let

$$\nabla_x I_a(x, y) = \frac{I_a(x+1, y) - I_a(x-1, y)}{2} \quad \text{and} \quad \nabla_y I_a(x, y) = \frac{I_a(x, y+1) - I_a(x, y-1)}{2}, \quad (10)$$

the gradient $\nabla I_a(x, y)$ is then given by

$$\nabla I_a(x, y) \approx \sqrt{[\nabla_x I_a(x, y)]^2 + [\nabla_y I_a(x, y)]^2}. \quad (11)$$

Fig. 3 shows a one-dimensional example, with real data, to illustrate the effect of TV regularization. This pattern (in green) includes three tightly-packed areas and a relatively isolated one. Without TV regularization, the aerial image intensity (in red) plotted in (a) has many locations with substantial signals where there should be no pattern; consequently, the error margin with the threshold is small. If the threshold reduces somewhat, we would observe spurious areas in the resulting binarized image. We can compare this with (b), where, with TV regularization, such “noise” is significantly reduced.

Nevertheless, such TV regularization also comes with some drawbacks. In this example, the signal content of the four features is also reduced, and thus also compromising the robustness of the resulting design, because if the threshold is increased, some features may be lost. In other words, the contrast of the aerial image is reduced. To ameliorate this, we design a weight matrix $W(x, y)$ that mediates the TV function. It takes small values around the transition areas, and large values elsewhere. Mathematically, the transition areas are given by our design pattern, $\bar{I}(x, y)$. We extract its edge by morphology, where the result, $E(x, y)$, is given by

$$E(x, y) = [\bar{I}(x, y) \oplus S(x, y)] - [\bar{I}(x, y) \ominus S(x, y)]. \quad (12)$$

Here, \oplus and \ominus denote dilation and erosion, respectively, and $S(x, y)$ is a 3×3 structure element with all one's. Morphological dilation expands the shape of the input binary image ($\bar{I}(x, y)$)

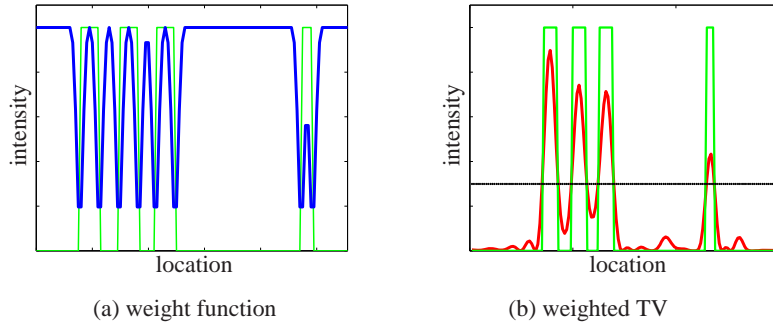


Fig. 4. The weight function $W(x,y)$ for TV regularization and its effect. The blue curve in (a) plots the weight function, and the red curve in (b) plots the aerial image intensity by using this weighted TV regularization. The green curve denotes the target design in both figures.

in this case), while erosion functions in the opposite way. These two operations are common approaches for binary image boundary extraction. Detailed mathematical description of them can be found in [26]. The weight matrix $W(x,y)$ is then given by

$$W(x,y) = [1 - E(x,y)] * G(x,y), \quad (13)$$

where $G(x,y)$ is a blurring function. Consequently, the first regularization term in Eq. (6) is then given by

$$\mathcal{R}_{\text{TV}}\{I_a(x,y)\} = \sum_{x,y} W(x,y) \cdot |\nabla I_a(x,y)|. \quad (14)$$

In our experiments we let $G(x,y)$ to be a 5×5 Gaussian kernel. This allows a smooth transition from penalty to no penalty, and vice versa. Referring to the earlier example, the weight function is given in Fig. 4(a). In (b) the resulting aerial image intensity curve by applying this weighted TV regularization is shown. The background intensity is smoothed similarly as in Fig. 3(b), while the aerial image contrast at transitions is better preserved.

3.3. Aerial image and contrast

With the above regularization, how are we going to control the area near the transition areas? In addition, what should be the control? The answer to the first question is straightforward, because the opposite of the above weight function, i.e., $1 - W(x,y)$, allows us to put the emphasis on the transition areas. The answer to the second question comes in two expressions.

First, our goal is to push intensity values away from the threshold t as far as possible. At places where the design pattern $\bar{I}(x,y) = 0$, we would like $I_a(x,y) \approx 0$; at places where $\bar{I}(x,y) = 1$, we would like $I_a(x,y) \approx 2t$, so the threshold would be mid-way. This is depicted in Fig. 5. We set the threshold mid-way such that the intensity on each side of the nominal threshold could be equally regularized to reduce the contour sensitivity to both higher and lower dose changes.

We can consolidate the two requirements in enforcing $I_a(x,y) \approx 2t\bar{I}(x,y)$. Thus, the second regularization term in Eq. (6), which can also be viewed as a penalty term, is given by

$$\mathcal{R}_{\text{aerial}}\{I_a(x,y), \bar{I}(x,y)\} = \sum_{x,y} [1 - W(x,y)] \cdot [I_a(x,y) - 2t\bar{I}(x,y)]^2. \quad (15)$$

Second, we would like the intensity slope to be as sharp as possible, since it is closely related to exposure latitude [1, p. 61]. The first derivative, which measures the rate of intensity change,

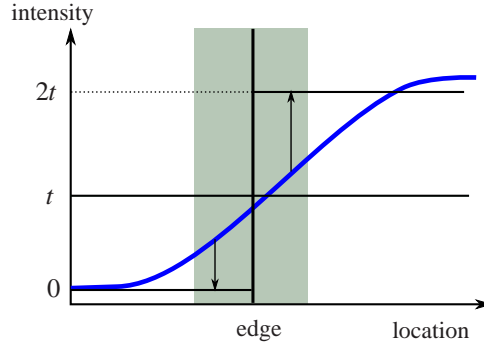


Fig. 5. Objective of aerial image intensity regularization. The shaded area denotes the vicinity of the target edge.

is the proper quantity to define the slope. A larger magnitude of the first derivative indicates a sharper image slope, or higher image contrast, and vice versa. The ideal target binary image has a theoretically infinite first derivative at its edge. To increase the exposure latitude, we force $\nabla_x I_a(x, y)$ and $\nabla_y I_a(x, y)$ to be close to $\nabla_x \bar{I}(x, y)$ and $\nabla_y \bar{I}(x, y)$, respectively. This is because most circuit designs use manhattan-shape features, which contain vertical and horizontal edges only. Thus the third regularization term $\mathcal{R}_{\text{contrast}}$ in Eq. (6) is

$$\mathcal{R}_{\text{contrast}}\{I_a(x, y), \bar{I}(x, y)\} = \sum_{x, y} [1 - W(x, y)] \cdot \left\{ \left| \nabla_x [I_a(x, y) - \bar{I}(x, y)] \right| + \left| \nabla_y [I_a(x, y) - \bar{I}(x, y)] \right| \right\}. \quad (16)$$

4. Statistical model for process robustness

Our discussion thus far is based on the assumption of an ideal imaging system without any process error [27]. We extend this optimization framework to a robust model by explicitly incorporating process variations, namely, dose variation and focus variation. As a reasonable assumption, we consider the process variations as independent, normally distributed random variables [28]. Specifically, dose variation can be accounted for by varying the threshold t ; focus variation, parameterized by β , is modeled by adding a phase term to the optical transfer function as

$$\tilde{H}(f, g; \beta) = \hat{H}(f, g) \cdot e^{i\pi\beta \frac{\text{NA}^2}{\lambda} (f^2 + g^2)}, \quad (17)$$

where NA is the numerical aperture, λ is the incident light wavelength, and $\beta\text{NA}^2/\lambda$ gives a normalized defocus quantity. A detailed mathematical description of the defocus model can be found in Ref. [10].

To compute solutions that are robust to process variations, the average wafer performance is optimized by minimizing the expectation of \mathcal{D} with respect to dose and focus fluctuations. The problem to be solved is thus described by a statistical model as

$$\begin{aligned} & \text{minimize} && \text{E}\{\mathcal{D}\{I(x, y), \bar{I}(x, y)\}\} \\ & \text{subject to} && M(x, y) \in \{0, 1\} \\ & && \hat{J}(f, g) \geq 0, \end{aligned} \quad (18)$$

where $\text{E}\{\cdot\}$ takes the expectation operation over t and β . However the expectation integral is difficult to compute due to the nonlinearity of \mathcal{D} . To tackle this problem, we discretize t to take

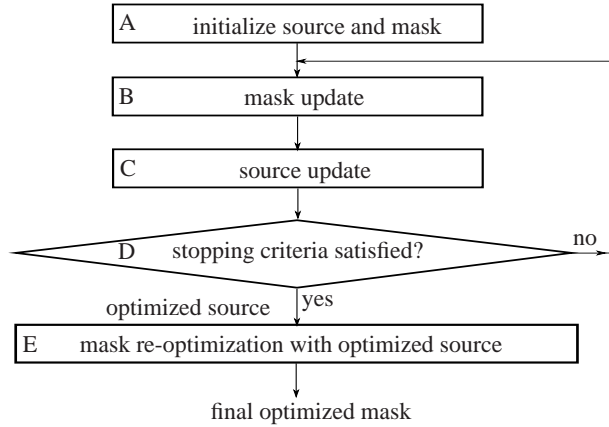


Fig. 6. Optimization procedure of SMO.

on a set of values t_m with probability $p(t_m)$, and discretize β to take on a set of values β_n with probabilities $p(\beta_n)$. The function $E\{\mathcal{D}\{I(x,y), \bar{I}(x,y)\}\}$ is computed as

$$E\{\mathcal{D}\{I(x,y), \bar{I}(x,y)\}\} \approx \sum_{m,n} p(t_m)p(\beta_n)\{\mathcal{D}\{I(x,y;t_m,\beta_n), \bar{I}(x,y)\}\}. \quad (19)$$

5. Optimization procedure

Given the cost function in Eq. (18), we minimize it by iteratively updating the source function and the mask pattern. The optimization procedure consists of multiple functional blocks as shown in the flow diagram in Fig. 6 with labels A to E.

In block A, we initialize the source and the mask. For the former we choose a traditional annular illumination; for the latter, the target design is the natural choice. Blocks B to D then form the core of the optimization process. First, the mask is updated by fixing the source (block B); then, the source is updated by fixing the mask (block C). This generates a new source-mask pair. Block D then checks if a pre-defined stopping criterion is met. (The simplest criterion can be a fixed number of steps, which is what we adopt for the simulations in the next section, but we can also use the value of the objective function \mathcal{D} as an indication of when to stop the iterations.) Blocks B to D are run again if it is not. Otherwise, we perform a final mask optimization to go along with the source illumination.

Below we explain in details how the mask and source updates are performed. We compute the updates using the nonlinear Hestenes-Stiefel conjugate gradient method [29, p. 123]; as such, each update consists of n iterative steps. We use a superscript with brackets to denote the current step. Also, we omit the designation (x,y) for brevity when no confusion arises.

5.1. Mask update

With the approximate objective function in Eq. (19), we first compute the gradient of $\mathcal{D}\{I(x,y;t_m,\beta_n), \bar{I}(x,y)\}$, denoted as $\nabla_M \mathcal{D}_{m,n}^{(0)}$. The derivation is found in the Appendix. We then sum it for all values of m and n to obtain the gradient of $E\{\mathcal{D}\{I(x,y), \bar{I}(x,y)\}\}$, denoted by $\nabla_M E^{(0)}$, as

$$\nabla_M E^{(0)} = \sum_{m,n} p(t_m)p(\beta_n)\nabla_M \mathcal{D}_{m,n}^{(0)}. \quad (20)$$

We call this the initial mask update, and assign $q_M^{(0)} = \nabla_M E^{(0)}$.

The rest of the iterations is computed as follows. Assume we have completed the k th step. The $(k+1)$ th mask update $q_M^{(k+1)}$ is calculated by the following steps:

1. Compute the current mask using the last update $q_M^{(k)}$ by

$$M^{(k+1)} = M^{(k)} + \varepsilon q_M^{(k)}, \quad (21)$$

where ε is a small constant known as the step size, and is determined empirically in this work.

2. Calculate the gradient of $\mathcal{D}\{I(x, y; t_m, \beta_n), \bar{I}(x, y)\}$ with respect to $M^{(k+1)}$, which we denote as $\nabla_M \mathcal{D}_{m,n}^{(k+1)}$, in a way similar to the initial step. Then, as in Eq. (20), we compute

$$\nabla_M E^{(k+1)} = \sum_{m,n} p(t_m) p(\beta_n) \nabla_M \mathcal{D}_{m,n}^{(k+1)}. \quad (22)$$

3. Compute an update parameter $\theta_M^{(k+1)}$, given by

$$\theta_M^{(k+1)} = \frac{\sum_{x,y} [\nabla_M E^{(k+1)} \cdot (\nabla_M E^{(k+1)} - \nabla_M E^{(k)})]}{\sum_{x,y} [q_M^{(k)} \cdot (\nabla_M E^{(k+1)} - \nabla_M E^{(k)})]}, \quad (23)$$

which correlates the gradient and the previous mask update to the current gradient.

4. Compute $q_M^{(k+1)}$ by

$$q_M^{(k+1)} = -\nabla_M E^{(k+1)} + \theta_M^{(k+1)} q_M^{(k)}. \quad (24)$$

5.2. Source update

The source can be updated with a similar procedure. Note that the effective source $\hat{J}(f, g)$ is a normalized quantity by its total energy [1, p. 45], i.e.,

$$\hat{J}(f, g) = \frac{\hat{J}'(f, g)}{\sum_{f,g} \hat{J}'(f, g)}. \quad (25)$$

In the following, we use $\hat{J}'(f, g)$ in the updates.

We first compute the gradient of $\mathcal{D}\{I(x, y; t_m, \beta_n), \bar{I}(x, y)\}$ with respect to the source, denoted as $\nabla_{\hat{J}'} \mathcal{D}_{m,n}^{(0)}$. The derivation is again detailed in the Appendix. The initial source update is then

$$q_{\hat{J}'}^{(0)} = \nabla_{\hat{J}'} E^{(0)} = \sum_{m,n} p(t_m) p(\beta_n) \nabla_{\hat{J}'} \mathcal{D}_{m,n}^{(0)}. \quad (26)$$

The subsequent iterations take the following steps:

1. Compute the current source using the last update $q_{\hat{J}'}^{(k)}$ by

$$\hat{J}'^{(k+1)} = \hat{J}'^{(k)} + \varphi q_{\hat{J}'}^{(k)}, \quad (27)$$

where φ is again a small constant representing the step size.

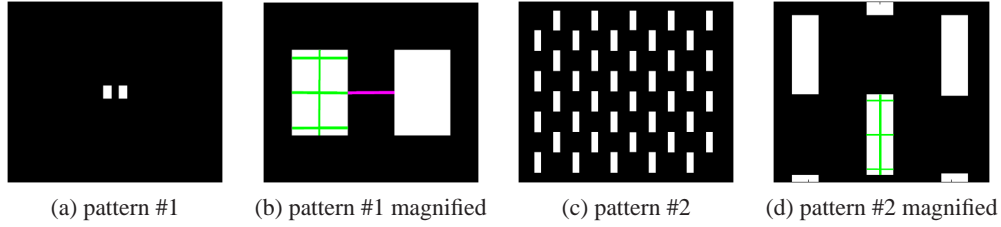


Fig. 7. Test patterns and critical locations for process window calculation. Color lines mark critical locations of the two test patterns.

2. Calculate the gradient of $\mathcal{D}\{I(x, y; t_m, \beta_n), \bar{I}(x, y)\}$ with respect to $\hat{j}^{(k+1)}$, i.e., $\nabla_{\hat{j}} \mathcal{D}_{m,n}^{(k+1)}$, and

$$\nabla_{\hat{j}} E^{(k+1)} = \sum_{m,n} p(t_m) p(\beta_n) \nabla_{\hat{j}} \mathcal{D}_{m,n}^{(k+1)}. \quad (28)$$

3. Compute an update parameter $\theta_{\hat{j}}^{(k+1)}$, given by

$$\theta_{\hat{j}}^{(k+1)} = \frac{\sum_{f,g} [\nabla_{\hat{j}} E^{(k+1)} \cdot (\nabla_{\hat{j}} E^{(k+1)} - \nabla_{\hat{j}} E^{(k)})]}{\sum_{f,g} [q_{\hat{j}}^{(k)} \cdot (\nabla_{\hat{j}} E^{(k+1)} - \nabla_{\hat{j}} E^{(k)})]}. \quad (29)$$

4. Compute $q_{\hat{j}}^{(k+1)}$ by

$$q_{\hat{j}}^{(k+1)} = -\nabla_{\hat{j}} E^{(k+1)} + \theta_{\hat{j}}^{(k+1)} q_{\hat{j}}^{(k)}. \quad (30)$$

During the source update, symmetry is important to avoid pattern placement error [15]. Usually a four-fold symmetry is imposed. To meet this specification, we force the gradient with respect to the source to be four-fold symmetric by averaging its four quarters' components [30].

6. Results

Here, we demonstrate the robust SMO algorithm in two distinct test patterns. First is a sparse pattern consisting of two rectangle shapes, as shown in Fig. 7(a). It is represented by a 151×151 matrix with a resolution of $10 \text{ nm} \times 10 \text{ nm}$ per pixel. Second is a dense poly pattern shown in Fig. 7(c), represented by a 473×473 matrix with a finer grid of $4 \text{ nm} \times 4 \text{ nm}$ per pixel. The imaging system parameters are set to be $\lambda = 193 \text{ nm}$ and $\text{NA} = 1.35$.

We compare the performance of SMO with that of mask optimization under a reference annular source. Certainly, with greater flexibility, the former should deliver better results than the latter; our objective here is to quantify how it is better, particular when it pertains to robustness. To do so, for each pattern we measure the feature size at a few critical locations, and then compute the process window according to the measured data. These locations include the main properties (width and length) of a feature, line-ends that are difficult to print, and the minimum feature size such as the space between two rectangles. They are marked by color lines (green if it is inside a feature, pink if outside) in Fig. 7(b) and (d) for patterns #1 and #2, respectively.

The size of the process window can be quantitatively measured by two parameters: exposure latitude (EL) and depth of focus (DOF). The former is the range of dose variation (% with respect to the nominal dose) where the feature size is within its tolerance, typically $\pm 10\%$ of its nominal size, at a certain defocus. The latter measures the largest acceptable defocus range

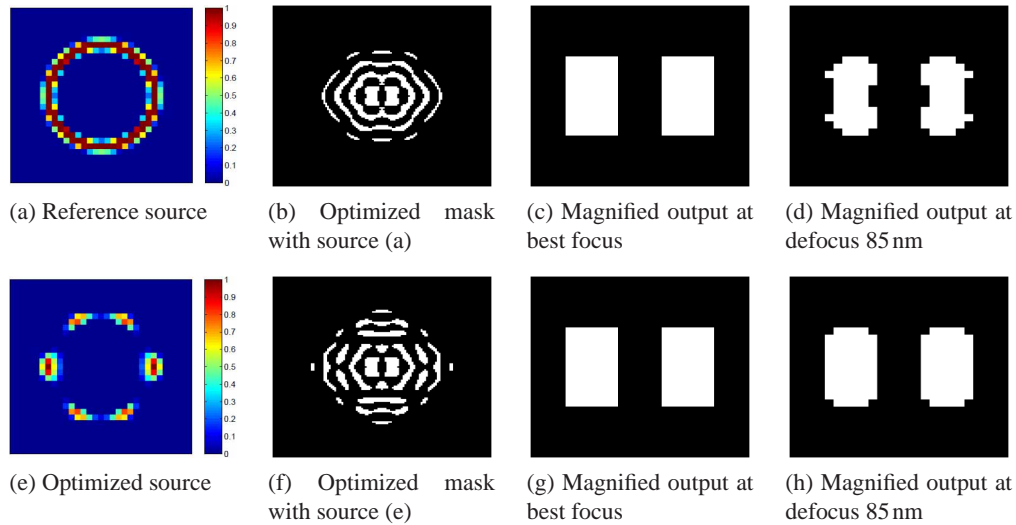


Fig. 8. Simulation results of test pattern #1.

(nm) under a fixed dose condition. Detailed descriptions of these quantities can be found in [1, p. 61–69]. The common method of measuring the process window is to examine how large EL or DOF can be when the other quantity is fixed. In the following we compare the DOFs by fixing the EL [31, 32]. Note that a larger DOF indicates a more robust performance.

6.1. A sparse pattern

In pattern #1, the two features are identical rectangles with height 110nm and width 60nm, separated by a 50nm space, which is the critical feature size for this pattern. We first assume that we have an annular source with its inner annulus $\sigma_{\text{inner}} = 0.7$ and outer annulus $\sigma_{\text{outer}} = 0.9$, as shown in Fig. 8(a). Note that we have applied a Gaussian blur on the annulus to mimic the reality, as a result of which the source does not take the same intensity inside the annulus.

We compute the corresponding optimized mask, together with its simulated output at best focus and at a defocus of 85nm, given in (b) to (d). In the second row, using the robust SMO algorithm presented in this paper, we show the resulting source and mask patterns in (e) and (f). The outputs at best focus and defocus are given in (g) and (h). In terms of pattern error, if we compare the results of mask optimization versus SMO, they give effectively identical output at best focus, but the latter delivers a pattern closer to the design that the former when there is defocus. As for the critical dimension (the 50nm space in between), the former results in a 20nm error at the center, while the latter keeps the nominal size. In other words, SMO gives a more robust design.

The optimized source has a strong component at the horizontal dipole location and four weak poles in the vertical direction. Since the small features in the target design are mainly along the horizontal direction, such a source configuration is more suitable than the circular reference source. We also calculate some numerical results. With a 10% EL, mask optimization with the reference source gains 150nm DOF, while SMO enlarges this number to 170nm for this pattern.

6.2. A dense pattern

We show the results of pattern #2 in Fig. 9 in a similar fashion, with a critical dimension of 60nm. The optimized source is given in (e), which is very different from that of the sparse pattern in the previous section. This source pattern is unlike any conventional illumination.

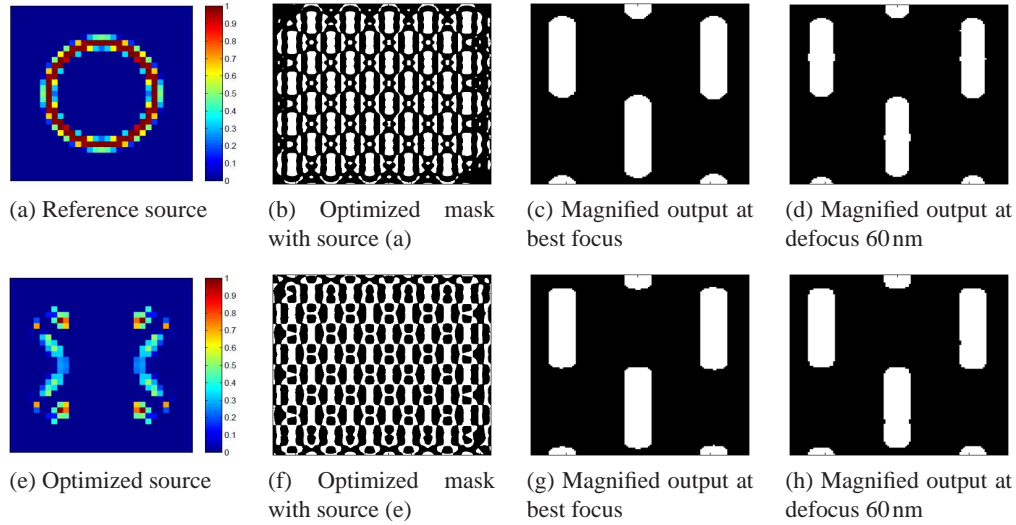


Fig. 9. Simulation results of test pattern #2.

Comparing the two wafer layouts printed at the nominal condition shown in Fig. 9 (c) and (g), we can observe that the poly line-ends are better printed in (g). When there is a 60 nm defocus, the wafer image in (h) still keeps the nominal feature width in general, though some local errors exist. But in (d), the focal change shrinks almost all polygons, resulting in a 8 nm linewidth error. Thus SMO delivers more robust wafer images than mask optimization only. Numerically, with a 5% EL, SMO increases the DOF from 78 nm to 128 nm.

7. Conclusion

In this paper, we propose a source-mask optimization method for process robustness enhancement. For this purpose, we introduce a cost function including not only the pattern fidelity term but also regularization terms to adjust the aerial image intensity distribution and its contrast. A statistical model is built by incorporating process variations explicitly into the optimization framework as random variables. Simulation results of sparse and dense patterns show conspicuous process window enlargement.

A. Appendix: Computing the gradients

Given the cost function $\mathcal{D}\{I(x,y;t_m,\beta_n),\bar{I}(x,y)\}$, we show here how to calculate its gradients with respect to any given mask pattern M and illumination source \hat{J}' . With a fixed t_m and β_n , we denote them as $\nabla_M \mathcal{D}$ and $\nabla_{\hat{J}'} \mathcal{D}$ respectively. As with Section 5, we omit the designation (x,y) for brevity.

From Eq. (6), we have

$$\nabla_M \mathcal{D} = \frac{\partial \mathcal{F}}{\partial M} + \gamma_1 \frac{\partial \mathcal{R}_{TV}}{\partial M} + \gamma_2 \frac{\partial \mathcal{R}_{aerial}}{\partial M} + \gamma_3 \frac{\partial \mathcal{R}_{contrast}}{\partial M} \quad (31)$$

$$\nabla_{\hat{J}'} \mathcal{D} = \frac{\partial \mathcal{F}}{\partial \hat{J}'} + \gamma_1 \frac{\partial \mathcal{R}_{TV}}{\partial \hat{J}'} + \gamma_2 \frac{\partial \mathcal{R}_{aerial}}{\partial \hat{J}'} + \gamma_3 \frac{\partial \mathcal{R}_{contrast}}{\partial \hat{J}'} \quad (32)$$

In the following we will present the analytical form of each term in Eq. (31) and (32). The derivation will be omitted when it is straightforward.

A.1. Computing $\nabla_M \mathcal{D}$

First we calculate the gradient of \mathcal{T} with respect to the mask M . In Eq. (2), $H(x, y; f, g)$ is replaced by the inverse Fourier transform of Eq. (17). For brevity we denote it as $\tilde{H}(f, g)$, and use $\tilde{H}'(f, g)$ to represent $\tilde{H}(-x, -y; f, g, \beta)$. Similar to calculating the gradient under coherent imaging system in [9], $\partial \mathcal{T} / \partial M$ is computed by

$$\begin{aligned} \frac{\partial \mathcal{T}}{\partial M} &= \frac{\partial \sum_{x,y} (I - \bar{I})^2}{\partial M} \\ &= \sum_{f,g} \hat{J}(f, g) \cdot \text{Re} \left\{ [2\alpha(I - \bar{I}) \cdot I \cdot (1 - I) \cdot (M * \tilde{H}(f, g))^{\dagger}] * \tilde{H}'(f, g) \right\}. \end{aligned} \quad (33)$$

Next we compute the gradient of \mathcal{R}_{TV} with respect to M . As in Eq. (10), the first derivatives are approximated by finite differences, and performed by matrix multiplication. Given an image U and a shifting matrix D , the first derivative $\nabla_x U$ is calculated by column operation UD , and $\nabla_y U$ by row operation DU . So the second term $\partial \mathcal{R}_{\text{TV}} / \partial M$ is given by

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{TV}}}{\partial M} &= \frac{\partial \sum_{x,y} W \cdot \sqrt{(DI_a)^2 + (I_a D)^2}}{\partial M} \\ &= \sum_{f,g} \hat{J}(f, g) \cdot \text{Re} \left\{ \left[\left(D^T \left(W \cdot [(DI_a)^2 + (I_a D)^2]^{-\frac{1}{2}} \cdot (DI_a) \right) \right. \right. \right. \\ &\quad \left. \left. \left. + \left(W \cdot [(DI_a)^2 + (I_a D)^2]^{-\frac{1}{2}} \cdot (I_a D) \right) D^T \right) \cdot (M * \tilde{H}(f, g))^{\dagger} \right] * \tilde{H}'(f, g) \right\}. \end{aligned} \quad (34)$$

We compute the third term $\partial \mathcal{R}_{\text{aerial}} / \partial M$ as

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{aerial}}}{\partial M} &= \frac{\partial \sum_{x,y} (1 - W) \cdot (I_a - 2t\bar{I})^2}{\partial M} \\ &= \sum_{f,g} \hat{J}(f, g) \cdot \text{Re} \left\{ [2(1 - W) \cdot (I_a - 2t\bar{I}) \cdot (M * \tilde{H}(f, g))^{\dagger}] * \tilde{H}'(f, g) \right\}. \end{aligned} \quad (35)$$

The fourth term $\partial \mathcal{R}_{\text{contrast}} / \partial M$ is calculated by

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{contrast}}}{\partial M} &= \frac{\partial \sum_{x,y} (1 - W) \cdot [|D(I_a - \bar{I})| + |(I_a - \bar{I})D|]}{\partial M} \\ &= \sum_{f,g} \hat{J}(f, g) \cdot \text{Re} \left\{ \left[\left(D^T \left((1 - W) \cdot \frac{D(I_a - \bar{I})}{\sqrt{[D(I_a - \bar{I})]^2}} \right) \right. \right. \right. \\ &\quad \left. \left. \left. + \left((1 - W) \cdot \frac{(I_a - \bar{I})D}{\sqrt{[(I_a - \bar{I})D]^2}} \right) D^T \right) \cdot (M * \tilde{H}(f, g))^{\dagger} \right] * \tilde{H}'(f, g) \right\}. \end{aligned} \quad (36)$$

A.2. Computing $\nabla_{\hat{J}} \mathcal{D}$

Given Eq. (32), we first compute the gradient of \mathcal{T} with respect to an arbitrary source point $\hat{J}'(f', g')$ as

$$\begin{aligned} \frac{\partial \mathcal{T}}{\partial \hat{J}'(f', g')} &= \frac{\partial \sum_{x,y} [I(x, y) - \bar{I}(x, y)]^2}{\partial \hat{J}'(f', g')} \\ &= \sum_{x,y} 2\alpha(I - \bar{I}) \cdot I \cdot (1 - I) \cdot \frac{|M * \tilde{H}(f', g')|^2 - I_a}{\sum_{f,g} \hat{J}'(f, g)}. \end{aligned} \quad (37)$$

The second term $\partial \mathcal{R}_{\text{TV}} / \partial \hat{J}'(f', g')$ in Eq. (32) is given by

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{TV}}}{\partial \hat{J}'(f', g')} &= \frac{\partial \sum_{x,y} W \cdot \sqrt{(DI_a)^2 + (I_a D)^2}}{\partial \hat{J}'(f', g')} \\ &= \sum_{x,y} W \cdot [(DI_a)^2 + (I_a D)^2]^{-\frac{1}{2}} \cdot \left[(DI_a) \cdot \frac{D \left[|M * \tilde{H}(f', g')|^2 - I_a \right]}{\sum_{f,g} \hat{J}'(f, g)} \right. \\ &\quad \left. + (I_a D) \cdot \frac{\left[|M * \tilde{H}(f', g')|^2 - I_a \right] D}{\sum_{f,g} \hat{J}'(f, g)} \right]. \end{aligned} \quad (38)$$

We compute the third term $\partial \mathcal{R}_{\text{aerial}} / \partial \hat{J}'(f', g')$ as

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{aerial}}}{\partial \hat{J}'(f', g')} &= \frac{\partial \sum_{x,y} (1-W) \cdot (I_a - 2t\bar{I})^2}{\partial \hat{J}'(f', g')} \\ &= \sum_{x,y} 2(1-W) \cdot (I_a - 2t\bar{I}) \cdot \frac{|M * \tilde{H}(f', g')|^2 - I_a}{\sum_{f,g} \hat{J}'(f, g)}. \end{aligned} \quad (39)$$

The fourth term $\partial \mathcal{R}_{\text{contrast}} / \partial \hat{J}'(f', g')$ is calculated by

$$\begin{aligned} \frac{\partial \mathcal{R}_{\text{contrast}}}{\partial \hat{J}'(f', g')} &= \frac{\partial \sum_{x,y} (1-W) \cdot \left[|D(I_a - \bar{I})| + |(I_a - \bar{I})D| \right]}{\partial \hat{J}'(f', g')} \\ &= \sum_{x,y} (1-W) \cdot \left[\frac{D(I_a - \bar{I})}{\sqrt{[D(I_a - \bar{I})]^2}} \cdot \frac{D \left[|M * \tilde{H}(f', g')|^2 - I_a \right]}{\sum_{f,g} \hat{J}'(f, g)} \right. \\ &\quad \left. + \frac{(I_a - \bar{I})D}{\sqrt{[(I_a - \bar{I})D]^2}} \cdot \frac{\left[|M * \tilde{H}(f', g')|^2 - I_a \right] D}{\sum_{f,g} \hat{J}'(f, g)} \right]. \end{aligned} \quad (40)$$

Acknowledgments

This work was supported in part by the University Research Committee of the University of Hong Kong under Project 10400898, by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Projects HKU 7134/08E, and by the UGC Areas of Excellence project Theory, Modeling, and Simulation of Emerging Electronics.