Basic approaches to simulation of resist mask formation in computational lithography

Nikita N. Balan, Vladimir V. Ivanov, Alexey K. Kuzovkov, Evgenia V. Sokolova, Evgeniy S. Shamin

1 Molecular Electronics Research Institute (JSC MERI), 12/1 1st Zapadnyi Proezd, Zelenograd, Moscow, 124460, Russia

Corresponding author: Nikita N. Balan (nbalan@niime.ru)

Received 26 December 2019 ♦ Accepted 18 February 2020 ♦ Published 30 March 2020

Citation: Balan NN, Ivanov VV, Kuzovkov AK, Sokolova EV, Shamin ES (2020) Basic approaches to simulation of resist mask formation in computational lithography. Modern Electronic Materials 6(1): 37–45. https://doi.org/10.3897/j.moem.6.1.55056

Abstract

Main currently used resist mask formation models and problems solved have been overviewed. Stages of “full physical simulation” have been briefly analyzed based on physicochemical principles for conventional diazonaphthoquinone (DNQ) photoresists and chemically enhanced ones. We have considered the concepts of the main currently used compact models predicting resist mask contours for full-scale product topologies, i.e., VT5 (Variable Threshold 5) and CM1 (Compact Model 1). Computation examples have been provided for full and compact resist mask formation models. Full resist mask formation simulation has allowed us to optimize the lithographic stack for a new process. Optimal thickness ratios have been found for the binary anti-reflecting layers used in water immersion lithography. VT5 compact model calibration has allowed us to solve the problem of optimal calibration structure sampling for maximal coverage of optical image parameters space while employing the minimal number of structures. This problem has been solved using cluster analysis. Clustering has been implemented using the $k$-means method. The optimum sampling is 300 to 350 structures, the rms error being 1.4 nm which is slightly greater than the process noise for 100 nm structures. The use of SEM contours for VT5 model calibration allows us to reduce the rms error to 1.18 nm for 40 structures.

Keywords

photoresist, projection photolithography, VT5, CM1, $k$-means clustering method.

1. Introduction

The resolution of projection photolithography (i.e., the minimum half-period size of printed structures) can be determined using the Rayleigh ratio [1]:

$$F = \frac{k_1 \lambda}{NA}$$

where $\lambda$ is the wavelength, $NA = n \sin \theta$ is the numerical aperture of the projection objective ($\theta$ is the aperture angle and $n$ is the refractive index of the medium) and $k_1$ is the coefficient determined by process parameters. In the classical Rayleigh formula for the resolution of optical systems this coefficient is 0.61 [2]. Various technics aimed at $k_1$ reduction are usually referred to in literature as resolution enhancement techniques (RET). These means are implemented using the so-called computational lithography methods [3] and include

- off-axis illumination [4];
- phase shifting reticles using [5–7];
- optical proximity correction [8].
Correct use of RET allows reducing $k_1$ to 0.28 or even smaller thus allowing printing of elements far smaller than the wavelength value.

The most important section of computational lithography is resist mask formation simulation. This simulation is always combined with one of the above listed $k_1$ coefficient reduction methods and may have the following implementations:

- so called “full physical simulation” based on physicochemical principles allowing computing a 3D resist mask configuration for relatively small wafer areas (units or decades of microns) (Fig. 1a);
- empirical approximation simulation allowing relatively rapid resist mask configuration computation (usually only the contour) for full-scale crystal topology with a number of simplified interpolation models, also called “compact” [9] (Fig. 1b). This type of simulation is in fact threshold image processing (with the complexity depending on the problem to be solved) of the computed optical image formed in the photoresist layer by the projection system of the lithographic tool. Image processing rules are determined from analysis of experimental data obtained for test structures printed on wafer (so called «model calibration»). One can also run a model experiment by «full» simulation of calibration structure printing.

This work provides an overview of the main currently used resist mask formation models with some examples of computations for existing and new processes.

2. “Full” simulation based on physicochemical principles

“Full” simulation describes all photomask formation stages (Fig. 2) – resistive films deposition, exposure, development, pre- and post-exposure heat treatments etc. (Fig. 2). This simulation is typically implemented through the so-called process-oriented CAD systems (KLA Proli-th, GenISys Lab, Panoramic Hyperlith etc.) used for photolithography processes simulation, e.g. for selection of optimal lithographic stack parameters, illuminator configuration etc. Full simulation is used in optical proximity correction procedure setup, SRAF rules development [10] and phase shift mask conditions definition [5, 6, 11, 12]. In the development of new photoresists blends this simulation allows predicting the dependence of the resist mask on the concentration of some components of the test photoresist composition.

The main process considered in the full simulation is latent image formation in the photoresist layer during exposure.

2D intensity distribution generated in wafer plane upon template illumination by an extended spatially incoherent radiation source (so-called “air” image) could be computed with the Hopkins method [1, 13]:

$$I(x) = \int \int TCC(f_x, f_y) \cdot T_\mathcal{N}(f_x) \cdot T_{\mathcal{R}}(f_y) \cdot e^{2\pi i f_x x - 2\pi i f_y y} \, df_x \, df_y,$$

where $TCC(f_x, f_y) = \int P(f_x + f'_x) P^*(f_y + f'_y) \tilde{S} (f'_x) \, df'_x$ is the so-called transmission cross-coefficient (TCC) depending only on the properties of the optical system and the illuminator shape,

$$P(f_x) = \begin{cases} 1, & f_x \leq \frac{N_A}{\lambda}, \\ 0, & f_x > \frac{N_A}{\lambda}, \end{cases}$$

is the pupil function, $\tilde{S} (f_x)$ is the normalized illuminator function and $T_\mathcal{N}(f_x) = F \{ E_x(x) \mathcal{N}(x) \}$ is the Fourier image of the product of the incident radiation amplitude and the template transmittance function (generally, it have complex value). For simplicity we consider the 1D case here.

The main advantage of the Hopkins method is that the transmission cross-coefficient TCC (a four-dimensional

Figure 1. Resist mask formation simulation examples: (a) 3D resist configuration computation using “full” simulation based on physical fundamentals (KLA Prolith CAD); (b) resist contour computation using compact models (Mentor Graphics Calibre CAD).
matrix in a discrete 2D case) only needs to be determined once following which one can compute intensity distributions for completely different source images using the already computed TCC thus achieving great time-saving.

The next stage is to determine the spatial light intensity distribution in the photoresist film. It can be computed using a number of approximations the simplest of which is intensity representation in the following form:

\[ I(x, z) = I_I(x)I_S(z), \]

where \( I_I(x) \) is the intensity of the “air” image determined e.g. with the Hopkins method and \( I_S(z) \) is the intensity distribution through film depth computed using the Fresnel coefficients [1, 2] in the assumption of normal flat wave incidence upon photoresist surface. This simplest approximation was used by F.H. Dill et al. in 1975 [14] for the derivation of the well-known equations describing the formation of latent images in photoresist layers. Intensity distribution through film depth is often oscillating due to the interference between the transmitted and reflected waves. The use of bottom anti-reflecting coating (BARC) the thickness and refractive index of which are matched with those of the photoresist film allows minimizing the intensity of the oscillations. Unfortunately this approximation is only acceptable for small digital apertures, whereas for \( NA > 0.35 \) one should use more complex and accurate models as described in detail in [1].

Then it is necessary to simulate the formation of a latent image in the photoresist layer during its exposure. To this end we use the Grothus-Draper Law (the First Law of Photochemistry) according to which only the absorbed portion of incident light can initiate a chemical reaction in the material. In turn, light absorption is described by the Bouguer-Lambert Law describing the attenuation of light intensity during radiation propagation in an absorbing medium:

\[ I(z) = I_0 \exp(-\alpha z), \]

where \( I_0 \) is the incident radiation intensity, \( z \) is the coordinate inside the material (with the origin on the photoresist film surface) and \( \alpha \) is the absorption coefficient [1].

It is important to understand that each component of the light-sensitive composition has its unique molar absorption ratio and that \( \alpha \) generally depends on the concentrations of the components. Typical positive photoresists intended for operation at 436 and 365 nm wavelengths contain novolac resin and a diazonaphthoquinone (DNQ) based photosensitive component (Fig. 3). Furthermore they contain organic solvents and photochemical reaction products (Fig. 4).

The transformation process of an optical image to a latent image (i.e., the photochemical reaction kinetics) is described by Dill’s set of differential equations [14]:

\[
\frac{\partial I(z, t)}{\partial z} = -I(z, t) \left[ AM(z, t) + B \right],
\]

\[
\frac{\partial I(z, t)}{\partial t} = -I(z, t) M(z, t) C,
\]
where $t$ is time, $M(z,t)$ is the relative concentration of the photosensitive component, $A$ is the exposure-dependent part of the absorption ratio, $B$ is the constant part of the absorption ratio and $C$ is the photosensitive component decomposition rate. The $A$, $B$ and $C$ parameters also referred to as Dill’s parameters are determined experimentally [15, 16] or [1]. Decomposition of the photosensitive component (Fig. 4) causes an abrupt increase in the photore sist film solubility in alkaline solutions.

Simulating DNQ photoresist exposure and development one should take into account the effects caused by preliminary heat treatment and the influence of residual solvent on reaction products diffusion during post-exposure bake. The main purpose of preliminary heat treatment (or post-coating heat treatment (Fig. 2)) is to initiate solvent evaporation and diffusion and partially decompose the photosensitive component. These phenomena eventually distort Dill’s parameters.

Conventional resists become inefficient at wavelengths of 248 nm or less due to strong light absorption by the novolac resin. Therefore the so-called chemically enhanced resists were developed in which photo acid generators (PAG) react with the incident radiation (Fig. 5). The acid released during this exposure modifies the polymer matrix during post-exposure bake and makes the polymer soluble (Fig. 6) while not being consumed during the reaction. Here, the catalysis principle is implemented: heating-induced acid diffuses inside the polymer and reacts with macromolecules, destroys protective hydrophobic groups, produces hydrophilic ones and then releases again.

$$D(M) = \exp(E_1 + E_2M + E_3M^2),$$

where $M(x,z)$ is the latent image or light-sensitive component concentration and $E_1$, $E_2$ and $E_3$ are experimentally determined coefficients.

The possibility of resist mask profile computation during full simulation allows predicting the generation of standing waves and obtaining the so-called “swing curves”, i.e., dependences of a parameter (e.g., critical size or full exposure dose) on photoresist mask thickness, as well as evaluating the so-called “process windows”, i.e., regions in the dose vs defocus coordinates in which the resist mask retains process-acceptable geometry.

### 3. Use of “compact” models in resist mask contour computation for full-scale product topology

The main objective of empirical approximation simulation in resist mask contour computation is usually to rapidly simulate a 2D resist mask geometry for full-scale lithographic layer topology during the topological correction of optical proximity distortions. As noted above this simulation is in fact threshold processing of a 2D air image computed using the Hopkins method.
The so-called design-oriented CAD systems implementing this simulation provide a tremendous time advantage in resist mask contour computation as compared with process-oriented ones [19] but their accuracy depends strongly on the quantity and quality of calibration data which could be obtained from experiments or process-oriented CAD simulation.

In some cases it is expedient to use compact models for process simulation. For example, they allow illuminator shape optimization [4] or rapid process window area evaluation [20] on the basis of a computed optical image only.

There are several types of compact models used in 2D resist contour computation for full-scale product topology. The most widely used are the VT5 (Variable Threshold 5) and CM1 (Compact Model 1) models, e.g. implemented in Mentor Graphics Calibre CAD.

The VT5 model has been developed from simpler variable threshold models, e.g. VTR and VTRE, and has the following operation principle. Reference marks are set at topological area contours for which image parameters are determined, e.g. maximum and minimum intensity along the mark, as well as the first and second derivatives of intensity by coordinate for the point the intensity at which is equal to the reference threshold. The photoresist edge position is determined through the threshold value \( T \) which in turn depends on the preset air image parameters \( T = T(I_{\text{max}}, I_{\text{min}}, dI/dx, d^2I/dx^2) \). The threshold function \( T \) is typically a first or second order polynomial. Photoresist edge position simulation accuracy within the model considered can be increased by using the result of photomask image convolution with a preset set of kernels and applying the threshold function to the resultant distribution. This approach allows for the effect of photoresist components diffusion and etching effects [21].

Model calibration implies determination of \( T \) polynomial coefficients while analyzing results of printing of the so-called test calibration matrix containing a set of simple structures (Fig. 7) providing for the maximal coverage of air image parameters space.

The more accurate CM1 model yields a contour formed by a cross-section at the constant threshold of the \( R(x,y) \) surface corresponding to the latent image in the resist [22]. The \( R(x,y) \) surface is determined by the following linear combination:

\[
R(x,y) = \sum_i c_i M_i(x,y),
\]

where \( c_i \) are the model calibration coefficients and \( M_i(x,y) \) are the terms obtained by optical image transformations \( R(x,y) \):
\[ M(x, y) = \left( \left( \nabla^k I_{x,y}(x, y) \right)^n \otimes G_s(x, y) \right)^{1/n}, \]

where \( G_s(x, y) \) is the Gauss-Laguerre function, the \( k, n \) and \( p \) parameters are user-defined for each term and the \( b \) and \( s \) parameters are computed during model calibration. The \( b \) parameter determines the linear-piecewise transformation of the \( I(x, y) \) function and characterizes the pattern of the interaction between acids and alkali in the photoresist layer, the \( s \) parameter being the diffusion length of active components.

The CM1 model provides good accuracy in comparison with VT5 for 65 nm or less processes. One reason is that CM1 indirectly uses simulation on the basis of physical fundamentals.

### 4. Lithographic stack layers thickness optimization

In this section we give an application example of “full” resist mask formation simulation on the basis of physical principles in lithographic stack layer thickness optimization for a new process.

When lithographic stack layers are not matched by their thickness and refractive indices, the overall stack reflection coefficient varies quasi-periodically in a range of 20–60 % depending on initial resist film thickness variation. Therefore the effective energy absorbed in the resist layer varies in approximately the same range and hence the critical structure size should also meet this limit. The curve describing this dependence is referred to as swing curve [1]. Varying the thicknesses and refractive indices of lithographic stack layers (especially in anti-reflecting layers) usually allows achieving the minimal swing function amplitude by minimizing the length of the swing curve:

\[ L = \int_{h_{\min}}^{h_{\max}} \sqrt{\left( \frac{dCD(h)}{dh} \right)^2} \, dh = \text{min}. \]

In our example we simulated the lithographic process at a 193 nm wavelength in an immersion medium having a refractive index of 1.44 with a phase attenuating photomask having a dark field transmittance coefficient of 0.06 and a phase rotation angle of 180 deg. The photomask was dark-field. The illuminator parameters were as follows: Quasar, Cross orientation, cut angle = 34 deg, \( \sigma_{ext} = 0.9 \) and \( \sigma_{int} = 0.8 \). We used a periodical linear array of 100 nm spaced 50 nm wide bright dashes. Figure 8a shows the CD swing curve and the 120 nm resist mask profile for a 21.398 mJ/cm² dose on non-optimized lithographic glass. It can be seen from Fig. 8 that the lithographic stack considered was not optimal.

Varying the thicknesses of the bottom anti-reflecting layers with simultaneous computation and analysis of the resultant CD swing curves allows selecting the optimal lithographic stack parameters.

![Figure 8: CD swing curves and resist profiles (120 nm height, 21.298 mJ/cm² dose): (a) non-optimized lithographic stack; (b) optimized lithographic stack.](image)

The best thicknesses of the bottom anti-reflecting layers were \( hB1 = 22 \) nm and \( hB2 = 26 \) nm. The respective stack CD swing curve and the 120 nm resist mask profile for a 21.398 mJ/cm² dose are shown in Fig. 8b. It can be seen that the resist profile did not contain standing waves.

### Table 1. Lithographic stack parameters.

<table>
<thead>
<tr>
<th>Lithographic stack</th>
<th>Notation</th>
<th>Not optimal</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersion</td>
<td>nl</td>
<td>1.44</td>
<td>1.44</td>
</tr>
<tr>
<td>BARC1</td>
<td>hB1, nm</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>nB1</td>
<td>1.9064</td>
<td>1.9064</td>
</tr>
<tr>
<td></td>
<td>kB1</td>
<td>0.6711</td>
<td>0.6711</td>
</tr>
<tr>
<td>BARC2</td>
<td>hB2, nm</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>nB2</td>
<td>1.7021</td>
<td>1.7021</td>
</tr>
<tr>
<td></td>
<td>kB2</td>
<td>0.196</td>
<td>0.196</td>
</tr>
<tr>
<td>Resist</td>
<td>hR, nm</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>nR</td>
<td>1.706</td>
<td>1.706</td>
</tr>
<tr>
<td></td>
<td>kR</td>
<td>0.00922</td>
<td>0.00922</td>
</tr>
<tr>
<td>Immersion protective coating</td>
<td>hI, nm</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>nl</td>
<td>1.53</td>
<td>1.53</td>
</tr>
</tbody>
</table>

This lithographic stack was considered acceptable. The stack layer thicknesses as well as the real and imaginary components of stack film materials refractive indices are summarized in the Table 1.

### 5. VT5 model calibration for optical proximity correction (OPC) procedure

The most time-consuming stage of compact model calibration process is measurement data acquisition. For
non-automated measurements, data acquisition may take dozens of hours and therefore OPC model development and the whole OPC solution become quite time-consuming. The necessity to reduce the number of measurements imposes the problem of optimal set of structures to be measured. As noted above the optimal sampling should maximally cover the optical image parameters space while employing the minimal number of structures.

The test calibration pad contained 3300 periodical linear structures sized 60 to 200 nm and spaced 140 to 1200 nm. The aim of the tests was to determine the optimal number of structures providing for an acceptable model accuracy the latter being evaluated by comparing simulation results with measurement data for 50 differently shaped structures that were not used for calibration. The accuracy criterion was chosen to be the rms error of simulation results.

The wavelength was $\lambda = 193$ nm, $NA = 0.75$. The illuminator parameters were as follows: Quasar, Cross orientation, $\sigma_{int} = 0.5$, $\sigma_{ext} = 0.8$ and cut angle = 30 deg (Fig. 9).

The tests were performed for different data arrays that included 40 to 360 structures with a step of 40. Each test structure was represented by a point in the image parameters space (Fig. 10a). When selecting points one should provide for the maximal coverage of the space occupied by the entire array of 3300 structures. This filtering can be effected using cluster analysis [23]: space points are grouped into clusters with one central point being chosen in each cluster (Fig. 10b). These central points represent the test structures to be included into the final set. In this work we performed clustering with the k-means method.

For each of the grouped data sets we calibrated the VT5 model with Mentor Graphics Calibre CAD and analyzed its accuracy. The results are presented in Fig. 11.

As we expected, sets containing 40 or 80 structures yielded a large error which decreased monotonically with the number of measurements.
an increase in the number of measurements. At above 280 structures the error stopped decreasing and reached an almost constant level of 1.4 nm. We therefore could conclude that the optimal calibration set is 300 to 350 test structures if sampling is performed using cluster analysis. The use of SEM contours for VT5 model calibration allows one to reduce the rms error to 1.18 nm for 40 structures. Simulation results are presented in Fig. 12.

6. Conclusion

This work describes two existing approaches to photomask formation simulation and demonstrates their applicability for projection photolithography and OPC recipe development. Full resist mask formation simulation allowed us to optimize the lithographic stack for a new process. VT5 model calibration with different data arrays showed that cluster analysis is a powerful sampling tool. The optimal sampling size made up at 300 to 350 structures.

References


