This paper was published in the proceedings of
Metrology, Inspection, and Process Control for Microlithography XIII,
SPIE Vol. 3677, pp. 415-434.

Data Analysis for Photolithography

Chris A. Mack, Sven Jug, Dale A. Legband

FINLE Technologies, Inc., P.O. Box 162712, Austin, TX 78716 USA
Voice: 512-327-3781, FAX: 512-327-1510, Email: chris_mack@finle.com

Abstract

This paper will propose standard methodologies for analyzing common lithographic data in three areas: photoresist contrast curves, swing curves, and focus-exposure matrices. For most data types, physics-based algebraic equations will be proposed to fit to the data. The coefficients of these equations will offer physical insight into the meaning and nature of the data. The equations will be fit to the data using standard non-linear least-squares fitting algorithms with standard statistical tests for removing data flyers and options for weighting the data. Analysis of the resulting curve fits will provide important information about the data. For the case of contrast curve data, the curve fits will yield resist contrast and dose-to-clear results. For swing curves, the swing ratio, period and the positions of the minimums and maximums will be provided. For focus-exposure data, process windows will be generated based on resist profile specifications. These process windows will then be analyzed by fitting rectangles or ellipses inside the window and determining the resulting exposure latitude/depth of focus trade-off. By specifying the desired exposure latitude, for example, the depth of focus and the best focus and best exposure to yield this maximum depth of focus will be calculated. Multiple process window overlaps can also be analyzed.

I. Introduction

Although there has been previous work in the area of tools and techniques for lithographic data analysis [1-3], there exists today no standards, or even commonly accepted practices, for the analysis of lithographic data such as swing curves and critical dimension (CD) focus-exposure matrices. An informal survey of semiconductor manufacturers has shown that most lithography engineers perform either no analysis or rudimentary spreadsheet analysis of focus-exposure matrix data to determine best focus and exposure, and very few fabs analyze this data to determine process windows or to calculate depth of focus in a rigorous way. Even simple analysis chores, such as finding the maximum of a swing curve, is typically done by "eye-balling" a graph of the data rather than using mathematical/statistical techniques for assessing the data.

This paper will propose standard methodologies for analyzing common lithographic data in three areas: positive and negative resist contrast curves, reflectivity, E_o or CD swing curves, and focus-exposure matrices (using CD, sidewall angle, and/or resist loss data). For most data types, physics-based algebraic equations will be proposed to fit to the data. Wherever possible, the coefficients of these equations will offer physical insight into the meaning and nature of the data. The equations will be fit to the experimental data using standard non-linear least-squares fitting algorithms with standard statistical tests for removing data flyers and options for weighting the data. Analysis of the resulting curve fits will provide important information about the data. For the case of contrast curve data, the curve fit results will yield resist contrast and dose-to-clear. For swing curves, the swing ratio, period and the positions of the minimums and maximums will be provided. For focus-exposure data, process windows will be generated based on resist profile specifications of linewidth, sidewall angle, and/or resist loss. These process windows will then be analyzed by fitting rectangles or ellipses inside the window and determining the resulting exposure latitude/depth of focus trade-off. As an example, specifying the desired exposure latitude leads to a unique determination of the depth of focus and the best focus and best exposure to yield this maximum depth of focus. Multiple process window overlaps can also be analyzed.

After describing the techniques for analyzing experimental data, examples will be provided that show the value of these methods.

II. Photoresist Contrast Curves

The use of "contrast" to describe the response of a photosensitive material dates back over one hundred years. Hurter and Driffield measured the optical density of photographic negative plates as a function of exposure [4]. The "perfect negative" was one which exhibited a linear variation of optical density with the logarithm of exposure. A plot of optical density versus log-exposure showed that a good negative exhibited a wide "period of correct representation." Hurter and Driffield called the slope of this curve in the linear region γ, the "development constant." Negatives with high values of γ were later termed "high contrast" negatives because the photosensitive emulsion quickly changed from low to high optical density when exposed.

Photolithography evolved from photographic science and borrowed many of its concepts and terminology. When exposing a photographic plate, the goal is to change the optical density of the material. In lithography, the goal is to remove resist. Thus, an analogous Hurter-Driffield (H-D) curve for lithography might plot resist thickness after development versus log-exposure. Often, the initial thickness of the resist is normalized to one, so that the H-D curve displays the relative thickness remaining.

Following the definition of γ from Hurter and Driffield, the photoresist "contrast" has traditionally been defined as the slope of the lithographic H-D curve at the point where the thickness goes to zero. Thus,

$$\gamma = \pm \frac{1}{T} \frac{dT_r}{d \ln E}\bigg|_{E_o} \qquad (1)$$

where T_r is the resist thickness remaining after development, T is the resist thickness before development, E is the nominal exposure energy, and E_o is the energy at which T_r is just zero. E_o is called the clearing dose for positive photoresists and the gel dose for negative systems. The positive sign in equation (1) is used for negative resists and the minus sign is used for positive systems in order to keep the value of γ positive. (Note that often a base-10 logarithm is used in equation (1) rather than the natural logarithm. Here, the natural log will always be employed.) Figure 1 shows a typical example of a contrast (H-D) curve for a positive resist.

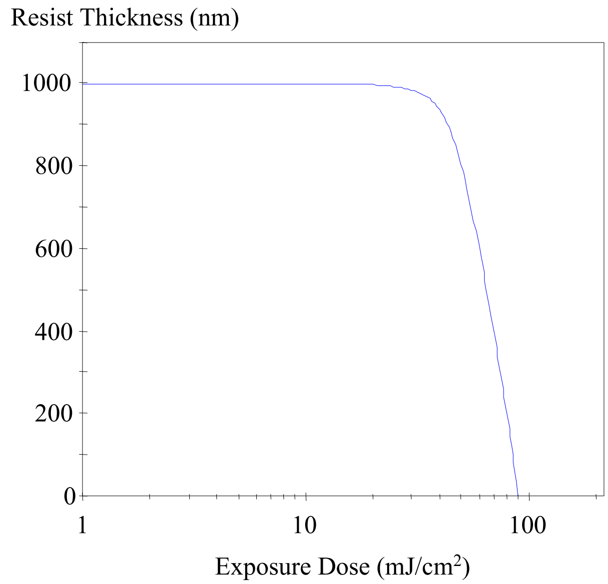


Figure 1. Typical photoresist contrast (H-D) curve for a positive photoresist.

There are many approaches to measuring the photoresist contrast [5]. For example, a somewhat dated ASTM standard specifies that the "linear" region of the curve between 0.1 and 0.7 relative resist thickness remaining should be fit with a straight line, the absolute value of the slope being defined as the contrast [6]. An extrapolation of this line to zero resist thickness defines the dose to clear (E_o). Another approach is to fit the entire resist thickness versus log-exposure curve and then apply equation (1) to the best-fit curve.

In order to derive an expression that adequately describes a typical H-D curve, the basic approach of Ziger and Mack [7] will be used. The thickness of resist remaining after an open frame exposure and development can be expressed as

$$T_r = T - \int_0^{t_{dev}} R\,dt \qquad (2)$$

where R is the development rate of the resist and t_{dev} is the development time. If the development rate does not vary appreciably with depth into the resist, then

$$T_r \approx T - R\,t_{dev} \qquad (3)$$

where R can now be thought of as an average of the development rate through the removed resist. The exposure dependence of R will determine the exposure dependence of the resist thickness remaining. One of the simplest models for development rate of a positive resist is the Mack model [8] with the parameter $m_{th} \ll 0$:

$$R = R_{max}\,(1-m)^n + R_{min} \qquad (4)$$

where R_{max} is the maximum development rate, R_{min} is the minimum development rate, n is the dissolution selectivity parameter, and m is the relative concentration of the compound which changes with exposure (for example, the diazonaphthoquinone). Further, the exposure kinetics relate m to the actual integrated exposure dose at some point in the film, E_z.

$$m = e^{-C E_z} \tag{5}$$

where C is the exposure rate constant. Although effects such as photoresist bleaching complicate the picture [9], it is a good approximation to say that the actual dose in the film is directly proportional to the incident dose E. Thus, equation (5) can be approximated as

$$m = e^{-E / E^*} \tag{6}$$

where E^* is a dose sensitivity term that is inversely proportional to the exposure rate constant.

Applying equations (4) and (6) to the thickness remaining expression of equation (3),

$$T_r = T_o - \Delta T_{max} \left(1 - e^{-E / E^*}\right)^n \tag{7}$$

where $T_o = T - R_{min}\, t_{dev}$ = resist thickness remaining for no exposure, and
$\Delta T_{max} = R_{max}\, t_{dev}$ = maximum possible resist loss (assuming a very thick resist).

Equation (7) is the final expression for the contrast curve relating the resist thickness remaining to exposure dose for a positive resist. The case of a negative resist can be easily handled by noting that the development rate equation (4) changes to

$$R = R_{max}\,(m)^n + R_{min} \tag{8}$$

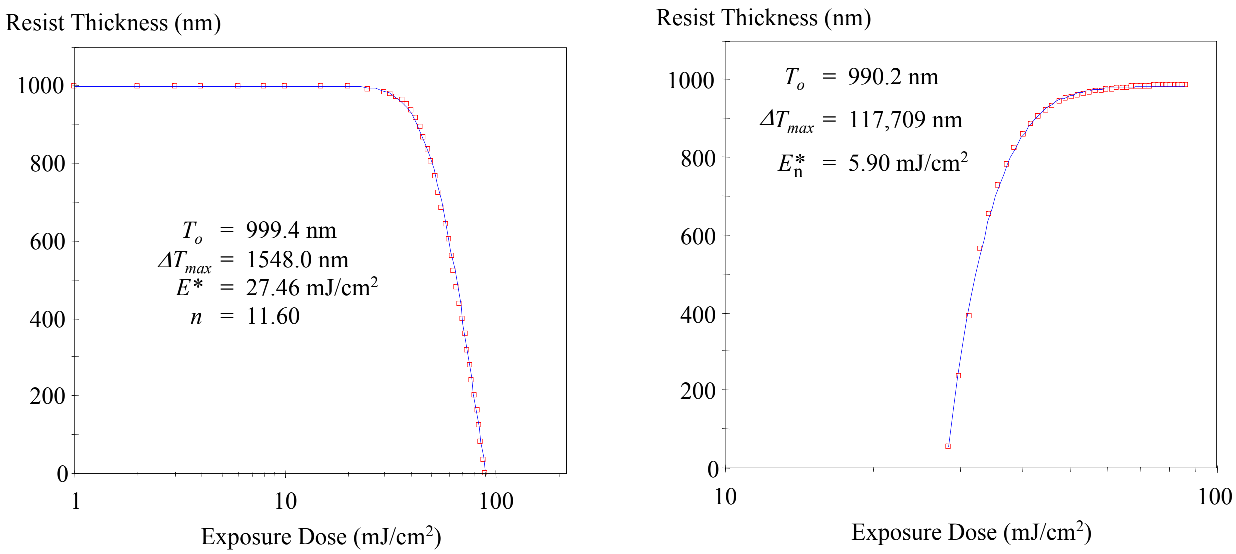
giving an equivalent final expression for the contrast curve

$$T_r = T_o - \Delta T_{max}\, e^{-nE / E^*} \tag{9}$$

Since both n and E^* are resist-dependent parameters, they can be lumped together into a new sensitivity term for negative resists, E_n^*

$$T_r = T_o - \Delta T_{max}\, e^{-E / E_n^*} \tag{10}$$

Examples of the applications of equations (7) and (10) are shown in Figure 2.



(a) (b)

Figure 2. Examples of curve fits to contrast curve data for (a) positive, and (b) negative resists.

From the curve fit equations, the contrast and the dose to clear can be extracted directly, making the (usually good) assumption that $T \approx T_o$.

Table I. Derived values of γ and E_o for the contrast (H-D) curve fit functions.

Positive Resist	***Negative resist***
$E_o = -E^* \ln\left(1-\left(\frac{T_o}{\Delta T_{max}}\right)^{1/n}\right)$	$E_o = -E_n^* \ln\left(\frac{T_o}{\Delta T_{max}}\right)$
$\gamma = \frac{E_o}{E^*} n\left(\left(\frac{\Delta T_{max}}{T_o}\right)^{1/n} - 1\right)$	$\gamma = \frac{E_o}{E_n^*}$

As an example, the data in Figure 2a gave a calculated value of $\gamma = 1.47$, whereas the ASTM linear fit between 10% and 70% of the resist thickness produced $\gamma = 1.39$.

III. Swing Curves

Exposing a photoresist involves the propagation of light through a thin film of partially absorbing material (the resist) coated on a substrate which is somewhat reflective. The resulting thin film interference effects include standing waves [10] and swing curves [11]. Generically, a swing curve is the sinusoidal variation of some lithographic parameter with resist thickness. There are several parameters which vary in this way, but the most important is the critical dimension (CD) of the photoresist feature being printed. Figure 3a shows a typical CD swing curve for *i*-line exposure of a 0.5 µm line on silicon.

The change in linewidth is quite large (more than the typical ±10% tolerance) for relatively small changes in resist thickness. Another swing curve is the E_o swing curve, showing the same sinusoidal swing in the photoresist dose-to-clear. For a resist thickness which requires a higher dose-to-clear, the photoresist will, as a consequence, require a higher dose to achieve the desired line size. But if the exposure dose is fixed (as it was for the CD swing curve), the result will be an underexposed line which prints too large. Thus, it follows that the E_o and CD swing curves result from the same effect. The final swing curve measures the reflectivity of the resist coated wafer as a function of resist thickness (Figure 3b). Although reflectivity is further removed from lithographic metrics such as E_o or CD, it is the reflectivity swing curve which provides the most insight as to the cause of the phenomenon.

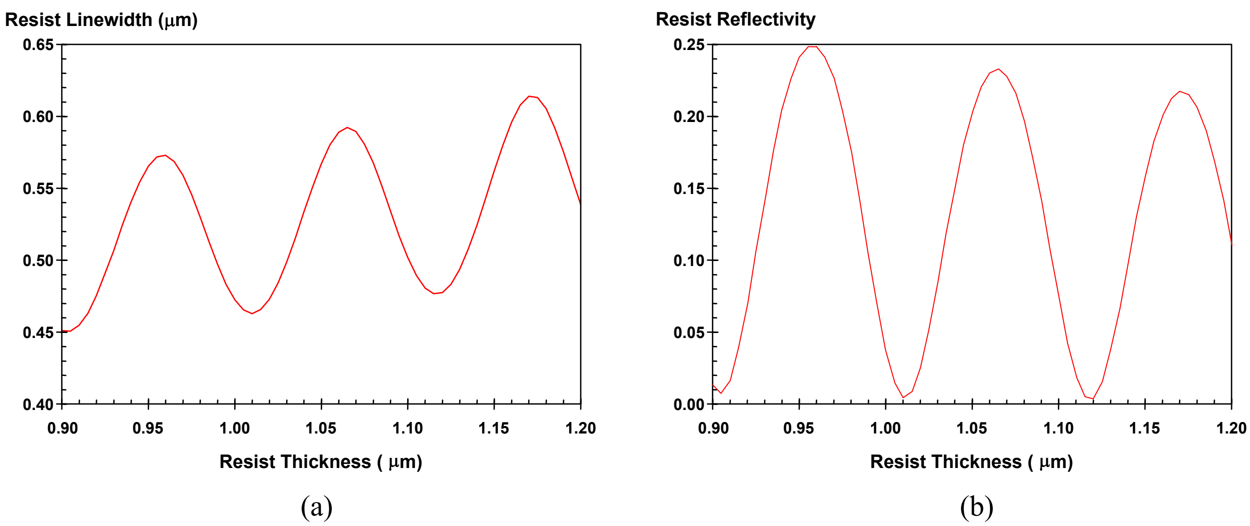


Figure 3. Swing curves showing a sinusoidal variation in (a) resist linewidth, and (b) reflectivity as a function of resist thickness.

The reflectivity swing curve shows that variations in resist thickness result in a sinusoidal variation in the reflectivity of the resist coated wafer. Since the definition of reflectivity is the total reflected light intensity divided by the total incident intensity, an increase in reflectivity results in more light which does not make it into the resist. Less light being coupled into the resist means that a higher dose is required to affect a certain chemical change in the resist, resulting in a larger E_o. Thus, the E_o and CD swing curves can both be explained by the reflectivity swing curve.

A. Reflectivity Swing Curve

What causes the reflectivity swing curve of Figure 3b? Of course, the answer lies in the thin film interference effects. Using the simple geometry shown in Figure 4a, a thin photoresist (layer 2) rests on a thick substrate (layer 3) in air (layer 1). Each material has optical properties governed by its complex index of refraction, $\boldsymbol{n} = n - i\kappa$. If we illuminate this film stack with a monochromatic plane wave normally incident on the resist, the total reflected light is made up of the incident beam reflecting off the air-resist interface and beams that have bounced off of the substrate and then were transmitted by the air-resist interface (Figure 4b). The total electric field reflection coefficient can be computed by totaling up all the reflected electric fields and then dividing by the incident field.

$$\rho_{total} = \frac{\rho_{12} + \rho_{23}\tau_D^2}{1 + \rho_{12}\rho_{23}\tau_D^2} \qquad (11)$$

where $\tau_{ij} = \dfrac{2\boldsymbol{n}_i}{\boldsymbol{n}_i + \boldsymbol{n}_j}$, the electric field transmittance from layer *i* to layer *j*,

$\rho_{ij} = \dfrac{\boldsymbol{n}_i - \boldsymbol{n}_j}{\boldsymbol{n}_i + \boldsymbol{n}_j}$, the electric field reflectance from layer *i* to layer *j*,

$\tau_D = e^{-i2\pi \boldsymbol{n}_2 D/\lambda}$, the internal transmittance of the resist film,

λ = the wavelength, and

D = the resist thickness.

The calculation of the intensity reflectivity (the square of the magnitude of equation (11)) is simplified by ignoring the imaginary part of ρ_{12}, giving

$$R = \frac{|\rho_{12}|^2 + |\rho_{23}|^2 e^{-\alpha 2D} + 2|\rho_{12}\rho_{23}|e^{-\alpha D}\cos(4\pi n_2 D/\lambda - \phi)}{1 + |\rho_{12}\rho_{23}|^2 e^{-\alpha 2D} + 2|\rho_{12}\rho_{23}|e^{-\alpha D}\cos(4\pi n_2 D/\lambda - \phi)} \qquad (12)$$

where ϕ is the phase of the complex reflection coefficient of the substrate, ρ_{23}, and where the absorption coefficient of the resist is related to the imaginary part of the resist index of refraction by

$$\alpha = \frac{4\pi\kappa_2}{\lambda}$$

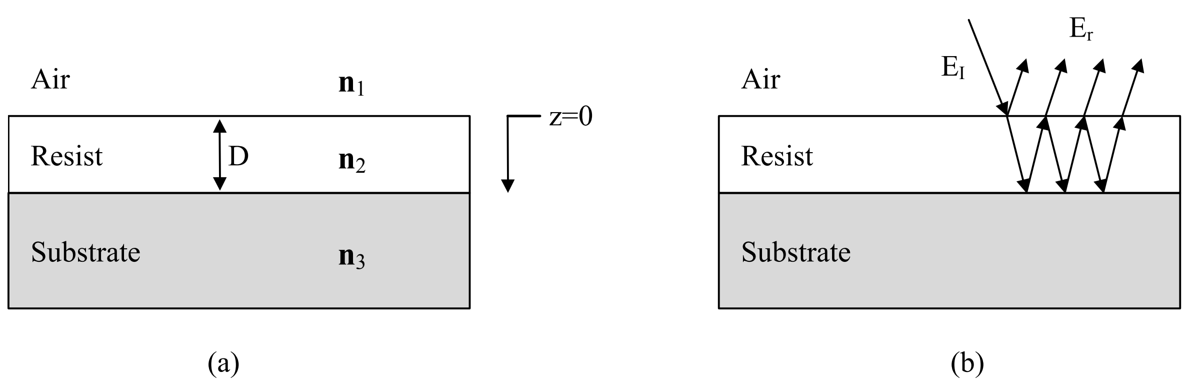


Figure 4. Film stack showing geometry for the swing curve derivation (oblique angles in (b) are shown for diagrammatical purposes only).

Equation (12) can be simplified when fitting experimental swing curve data if the data extends over only a few periods. Let D_o be the center of a range of resist thickness over which the swing curve expression will be approximated. Thus,

$$D = D_o + \Delta D \tag{13}$$

The resist absorption terms can then be approximated as linear with respect to the resist thickness when $\alpha \Delta D \ll 1$.

$$e^{-\alpha D} = e^{-\alpha D_o} e^{-\alpha \Delta D} \approx e^{-\alpha D_o}(1 - \alpha \Delta D) = e^{-\alpha D_o}(1 + \alpha D_o - \alpha D) \tag{14}$$

Thus, equation (12) will become

$$R \approx \frac{aD + b + (cD + d)\cos(2\pi D / P - \phi)}{1 + (cD + d)\cos(2\pi D / P - \phi)} \tag{15}$$

where $a = -2\alpha |\rho_{23}|^2 e^{-\alpha 2 D_o}$

$b = |\rho_{12}|^2 + (1 + 2\alpha D_o)|\rho_{23}|^2 e^{-\alpha 2 D_o}$

$c = -2\alpha |\rho_{12}\rho_{23}| e^{-\alpha D_o}$

$d = 2(1 + \alpha D_o)|\rho_{12}\rho_{23}| e^{-\alpha D_o}$

$P = \lambda / 2n_2$, the swing curve period,

and where the assumption is made that $|\rho_{12}\rho_{23}|^2 e^{-\alpha 2D} << 1$.

Often, the *(cD+d)* term is small compared to one, so that equation (15) simplifies further to

$$R \approx aD + b + (cD + d)\cos(2\pi D / P - \phi) \tag{16}$$

Both equations (15) and (16) can be used to fit experimental reflectivity swing curve data.

B. E_o and CD Swing Curves

An approximate behavior of the E_o swing curve can be obtained from the reflectivity results above. Since the amount of the light actually transmitted into the photoresist film is simply *1-R*, the energy deposited into the filmstack (E_{dep}) can be related to the incident dose (E_{inc}) by

$$E_{dep} = E_{inc}(1 - R) \tag{17}$$

Assuming that the energy deposited into the resist is linearly related to the energy deposited in the filmstack as a whole, the incident dose will equal E_o when the deposited energy reaches some critical dose, E_{crit}.

$$E_o = \frac{E_{crit}}{(1 - R)} \tag{18}$$

Algebraic manipulations and approximations similar to those used for the reflectivity swing curve will lead to an identical E_o swing curve form

$$E_o \approx aD + b + (cD + d)\cos(2\pi D / P - \phi) \quad (19)$$

where the numerical values of *a*, *b*, *c*, and *d* will differ from those defined in equation (15).

Likewise, the CD swing curve can be directly related to the reflectivity swing curve. If one approximates the CD versus deposited exposure dose curve to be linear over a small region near the nominal dose, equation (17) in combination with the reflectivity swing curve will yield

$$CD \approx aD + b + (cD + d)\cos(2\pi D / P - \phi) \quad (20)$$

where again the numerical values of *a*, *b*, *c*, and *d* will differ from previous values. Note, however, that *P* and ϕ, the period and phase offset of each of the three types of swing curves is the same.

Figure 5 shows an example of fitting equation (19) to typical E_o swing curve data taken for an i-line resist on bare silicon.

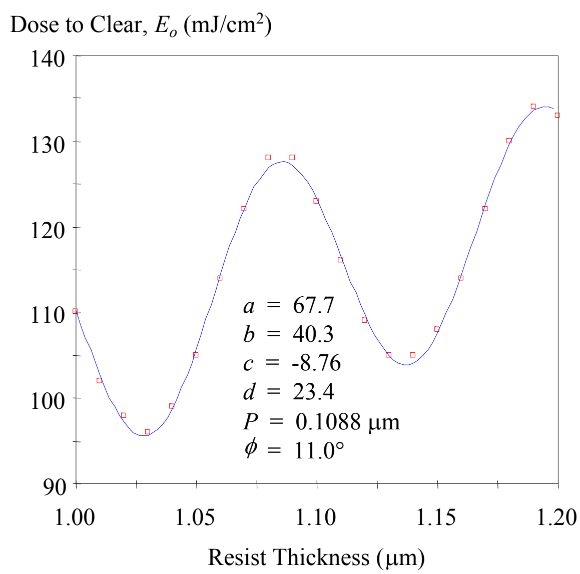


Figure 5. Best fit of equation (19) to E_o swing curve data.

Analysis of the fitted swing curve can lead to important information useful to lithographers. One of the reasons for measuring a swing is to determine the best resist thickness to use. Based on a number of considerations, either a swing curve minimum or a swing curve maximum is chosen (depending on whether CDs that are too big are more detrimental to yield and device performance than CDs that are too small). Then, the first swing curve maximum or minimum that exceeds some minimum resist thickness requirement is chosen as the optimum resist thickness. Using a fit of equation (16), (19), or (20) to experimental data is the best way to determine the optimum resist thickness to the highest degree of accuracy. It is interesting to note that the thicknesses that give the extrema of the cosine (i.e., $\cos\theta = \pm 1$) are not, in general, the thicknesses that produce the extrema of the swing curve.

Often, a simple measure of the magnitude of the variation caused by the swing curve is desired. Two metrics have been proposed, the *swing amplitude* and the *swing ratio*. The swing amplitude is simply the amplitude of the cosine term evaluated at a specific resist thickness.

$$Swing\ Amplitude\ =\ 2\,(cD+d)\times 100\% \qquad (21)$$

The swing ratio is a slightly more intuitive if somewhat less useful term. Taking the CD swing curve as an example, the linewidths of the first two maximums are averaged together to give CD_{max}. Then using the linewidth at the minimum between these two maximums, called CD_{min}, the swing ratio is defined as:

$$Swing\ Ratio\ =\ 2\,\frac{CD_{\max}-CD_{\min}}{CD_{\max}+CD_{\min}}\times 100\% \qquad (22)$$

IV. Focus-Exposure Matrix

Evaluating the effects of focus and exposure on the results of a projection lithography system (such as a stepper) is a critical part of understanding and controlling a lithographic process. This section will address the importance of focus by providing definitions of the *process window* and *depth of focus* (DOF) and applying these definitions to experimental focus-exposure data.

In general, DOF can be thought of as the range of focus errors that a process can tolerate and still give acceptable lithographic results. Of course, the key to a good definition of DOF is in defining what is meant by tolerable. A change in focus results in two major changes to the final lithographic result: the photoresist profile changes and the sensitivity of the process to other processing errors is increased. Typically, photoresist profiles are described using three parameters: the linewidth (or critical dimension, CD), the sidewall angle, and the final resist thickness. The variation of these parameters with focus can be readily determined for any given set of conditions. The second effect of defocus is significantly harder to quantify: as an image goes out of focus, the process becomes more sensitive to other processing errors such as exposure dose and develop time. Of these secondary process errors, the most important is exposure.

A. The CD FE Matrix

Since the effect of focus is dependent on exposure, the only way to judge the response of the process to focus is to simultaneously vary both focus and exposure in what is known as a *focus-exposure matrix*. Figure 6 shows a typical example of the output of a focus-exposure matrix using linewidth as the response (sidewall angle and resist loss can also be plotted in the same way) in what is called a Bossung plot [12]. As one can see, the shapes of the Bossung curves are quite complicated. As a result, most efforts to fit this data to an equation has involved the use of polynomials in focus (*F*) and exposure (*E*) [1-3]. One very general expression is

$$CD\ =\ \sum_{i=0}^{3}\sum_{j=0}^{4} a_{ij}\,E^{i}\,F^{j} \qquad (23)$$

Although this function has 20 adjustable coefficients, for most data sets good fits are obtained when a_{03}, a_{22}, a_{14}, a_{23}, a_{24}, a_{33}, and a_{34} are fixed and set to zero.

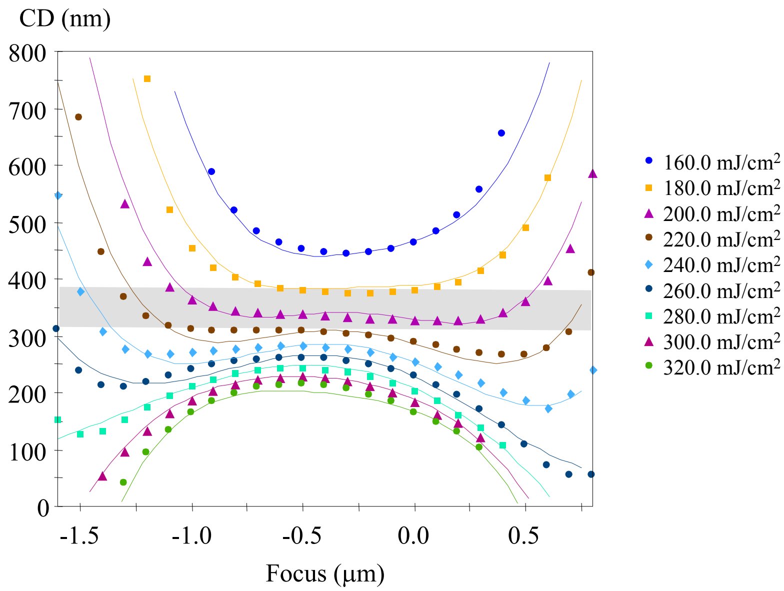


Figure 6. Example of the effect of focus and exposure on the resulting resist linewidth (symbols) and the best fit of this data (lines) to equation (23).

B. The Sidewall Angle FE Matrix

Sidewall angle data as a function of focus and exposure can be obtained from resist cross-sections. Although difficult to obtain, this data provides important information about the quality of the lithographic results. The following empirical equation has been derived to describe the behavior of sidewall angle (*SA*) as a function of focus (*F*) and exposure (*E*).

$$SA = \tan^{-1}\left(\left(1-\frac{E_o}{E}\right)^{\gamma}\left(\frac{E^*}{E}\right)^{\delta}\left(1+\left(\frac{F-F_o}{F^*}\right)^2\right)^{-1}\right), \quad F_o = aE+b \qquad (24)$$

where E_o = dose to clear-like term (units of exposure),
E^* = exposure sensitivity term (units of exposure),
γ = resist contrast-like term,
δ = strength of sidewall angle reduction at high doses,
F^* = depth of focus-like term (units of focus),
F_o = best focus-like term, (units of focus), which varies with exposure,
a = slope of exposure variation of best focus (units of focus/dose), and
b = constant term of exposure variation of best focus (units of focus).

C. The Resist Loss FE Matrix

Like sidewall angle, the loss of resist thickness in the center of a line feature can be measured using SEM cross-sections and provide insight into another mechanism for profile failure through focus and exposure. For positive resists, the following equation shows resist loss (*RL*) as a function of focus (*F*) and exposure (*E*).

$$RL = (RLS)(E)^n \left(1+\left(\frac{F-F_o}{F^*}\right)^2\right)^n + RL_{min} \qquad (25)$$

where RLS = resist loss sensitivity term (units of (resist loss)/(dose)^n),
n = resist contrast-like term,
F_o = best focus-like term (units of focus),
F^* = depth of focus-like term (units of focus), and
RL_{min} = minimum (unexposed) resist loss (units of resist loss).

For a negative resist, the equation becomes

$$RL = (RL_{max})\exp\left(-\frac{E}{E^*}\right)\left(1+\left(\frac{E}{E^*}\right)\left(\frac{F-F_o}{F^*}\right)^2\right) + RL_{min} \qquad (26)$$

where RL_{max} = maximum (unexposed) resist loss (units of resist loss), and
E^* = exposure sensitivity term (units of exposure dose).

An example of the use of equation (25) is shown in Figure 7.

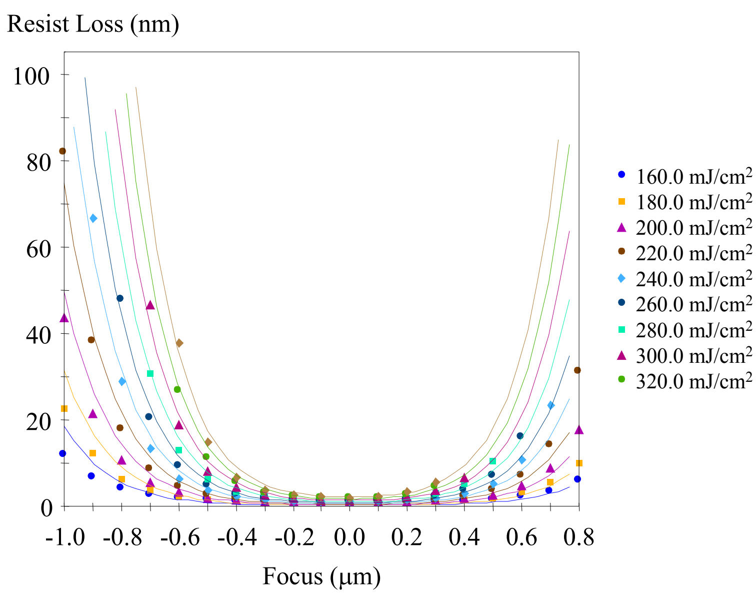


Figure 7. Example of the effect of focus and exposure on the resulting resist loss (symbols) and the best fit of this data (lines) to equation (25) for a positive resist.

D. Process Window

Of course, one output as a function of two inputs can be plotted in several different ways. For example, the Bossung curves could also be plotted as exposure latitude curves (linewidth versus exposure) for different focus settings. Probably the most useful way to plot this two-dimensional data set is a contour plot – contours of constant linewidth versus focus and exposure (Figure 8). By plotting the best fit equation (23) as a contour plot, smoothed data is automatically obtained, giving the best estimate of the true contour plot.

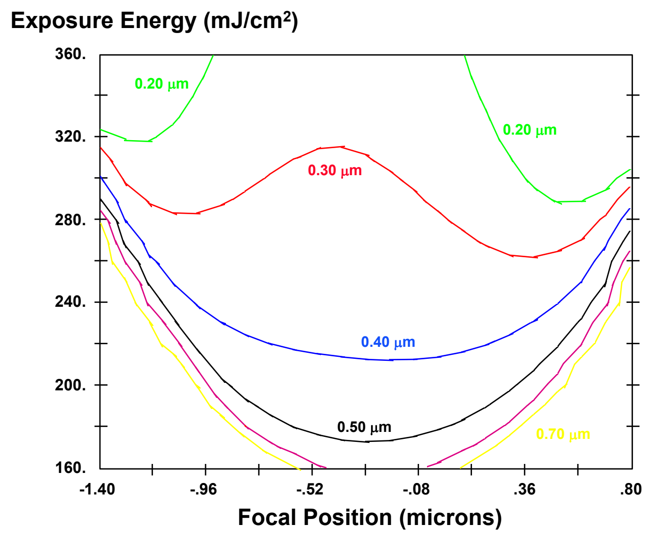


Figure 8. Displaying the fit to the data from a focus-exposure matrix in an alternate form, contours of constant CD versus focus and exposure.

The contour plot form of data visualization is especially useful for establishing the limits of exposure and focus that allow the final resist image to meet certain specifications. Rather than plotting all of the contours of constant CD, one could plot only the two CDs corresponding to the outer limits of acceptability – the CD specifications (Figure 9a). Because of the nature of a contour plot, other variables can also be plotted on the same graph. Figure 9b shows an example of plotting contours of CD (nominal ±10%), 80° sidewall angle, and 10% resist loss all on the same graph. The result is a *process window* – the region of focus and exposure that keeps the final resist profile within all three specifications.

The focus-exposure process window is one of the most important plots in lithography since it shows how exposure and focus work together to affect linewidth, sidewall angle, and resist loss. The process window can be thought of as a *process capability* – how the process responds to changes in focus and exposure. How can we determine if a given process capability is good enough? An analysis of the error sources for focus and exposure in a given process will give a *process requirement* [13]. If the process capability exceeds the process requirements, yield will be high. If, however, the process requirement is too large to fit inside the process capability, yield will suffer. A thorough analysis of the effects of exposure and focus on yield can be accomplished with yield modeling (for example, using ProCD) [14], but a simpler analysis can give useful insight and can be used to derive a number for depth of focus.

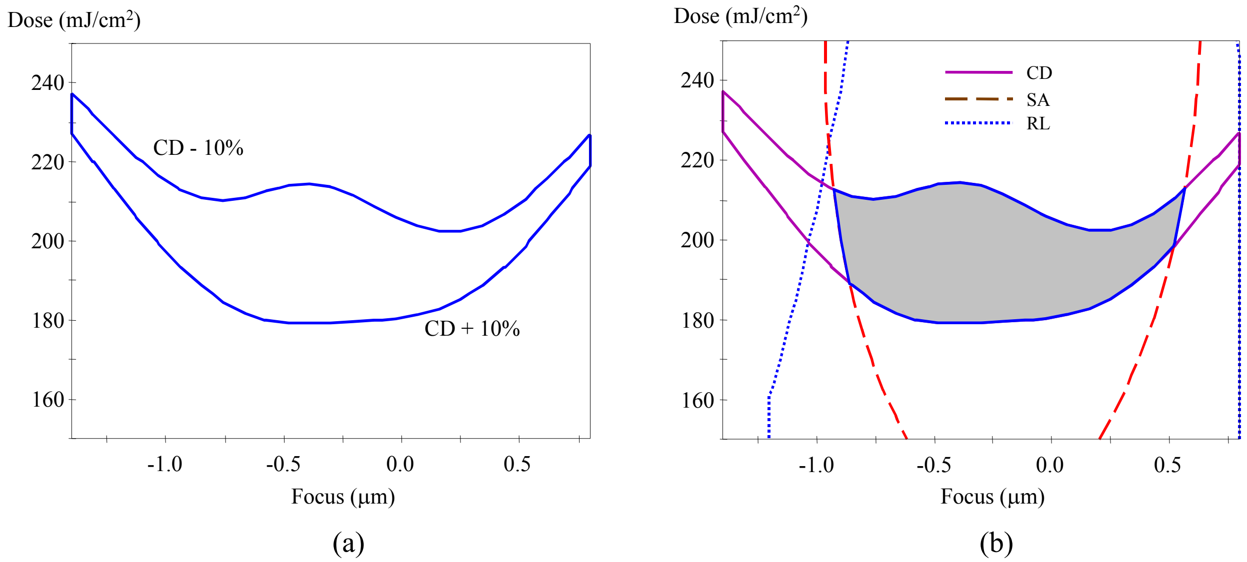


Figure 9. The focus-exposure process window is constructed from contours of the specifications for (a) linewidth, or (b) as an overlap of linewidth (CD), sidewall angle (SA), and resist loss (RL) specifications. The shaded area shows the overlap.

What is the maximum range of focus and exposure (that is, the maximum process requirement) that can fit inside the process window? A simple way to investigate this question is to graphically represent errors in focus and exposure as a rectangle on the same plot as the process window. The width of the rectangle represents the built-in focus errors of the processes, and the height represents the built-in dose errors. The problem then becomes one of finding the maximum rectangle that fits inside the process window. However, there is no one answer to this question. There are many possible rectangles of different widths and heights that are "maximum", i.e., they cannot be made larger in either direction without extending beyond the process window. (Note that the concept of a "maximum area" is meaningless here.) Each maximum rectangle represents one possible trade-off between tolerance to focus errors and tolerance to exposure errors. Larger DOF can be obtained if exposure errors are minimized. Likewise, exposure latitude can be improved if focus errors are small. The result is a very important trade-off between exposure latitude and DOF.

If all focus and exposure errors were systematic, then the proper graphical representation of those errors would be a rectangle. The width and height would represent the total ranges of the respective errors. If, however, the errors were randomly distributed, then a probability distribution function would be needed to describe them. For the completely random case, a Gaussian distribution with standard deviations in exposure and focus is used to describe the probability of a given error. In order to graphically represent the errors of focus and exposure, one should describe a surface of constant probability of occurrence. All errors in focus and exposure inside the surface would have a probability of occurring that is greater than the established cutoff. What is the shape of such a surface? For fixed systematic errors, the shape is a rectangle. For a Gaussian distribution, the surface is an ellipse. If one wishes to describe a "three-sigma" surface, the result would be an ellipse with major and minor axes equal to the three-sigma errors in focus and exposure.

Using either a rectangle for systematic errors or an ellipse for random errors, the size of the errors that can be tolerated for a given process window can be determined. Taking the rectangle as an example, one can find the maximum rectangle that will fit inside the processes window. Figure 10 shows an analysis of the process window where every maximum rectangle is determined and its height (the exposure latitude) plotted versus its width (depth of focus). Likewise, assuming random errors in focus and exposure, every maximum ellipse that fits inside the processes window can be determined. The horizontal width of the ellipse would represent a three-sigma error in focus, while the vertical height of the ellipse would give a three-sigma error in exposure. Plotting the height versus the width of all the maximum ellipses gives the second curve of exposure latitude versus DOF in Figure 10.

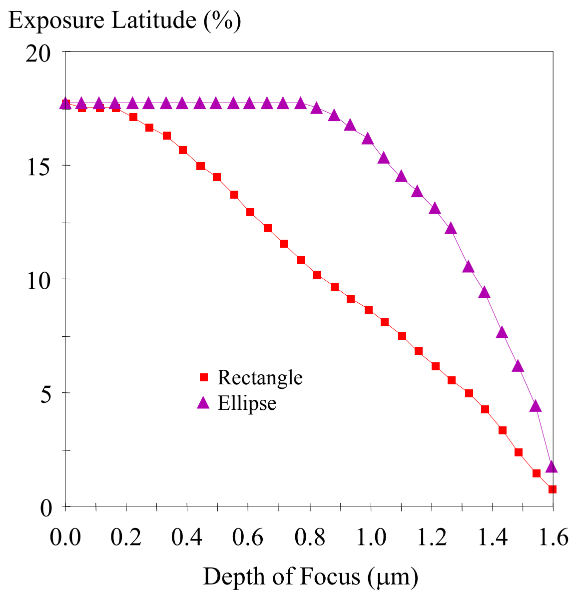


Figure 10. The process window of Figure 9 is analyzed by fitting all the maximum rectangles and all the maximum ellipses, then plotting their height (exposure latitude) versus their width (depth of focus).

The exposure latitude versus DOF curves of Figure 10 provide the most concise representation of the coupled effects of focus and exposure on the lithography process. Each point on the exposure latitude - DOF curve is one possible operating point for the process. The user must decide how to balance the trade-off between DOF and exposure latitude. One approach is to define a minimum acceptable exposure latitude, and then operate at this point; this has the effect of maximizing the DOF of the process. In fact, this approach allows for the definition of a single value for the DOF of a given feature for a given process. The depth of focus of a feature can be defined as *the range of focus that keeps the resist profile of a given feature within all specifications (linewidth, sidewall angle, and resist loss) over a specified exposure range*. For the example given in Figure 10, a minimum acceptable exposure latitude of 10%, in addition to the other profile specifications, would lead to the following depth of focus results:

DOF (rectangle) = 0.88 µm

DOF (ellipse) = 1.40 µm

As one might expect, systematic errors in focus and exposure are more problematic than random errors, leading to a smaller DOF.

The definition of depth of focus also leads naturally to the determination of best focus and best exposure. The DOF value read off from the exposure latitude versus DOF curve corresponds to one maximum rectangle or ellipse that fit inside the process window. The center of this rectangle or ellipse would then correspond to best focus and exposure for this desired operating point.

Overlapping process windows are used to find the ranges of focus and exposure that allow two or more different features to meet their respective profile specifications. For example, both dense and isolated features can be overlapped to find the depth of focus for simultaneously printing both features in spec. Process windows for horizontal and vertical features can be overlapped to show astigmatism, different feature sizes can be overlapped to show linearity, and process windows from many points in the field can show the "common corridor" depth of focus. Figure 11 shows a simple example of two overlapping CD process windows.

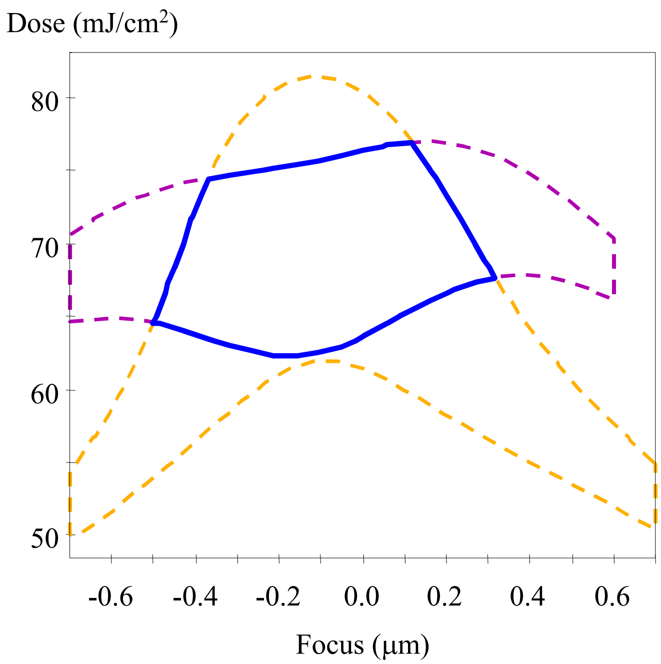


Figure 11. An example of overlapping process windows for dense and isolated lines.

V. Curve Fitting and Statistical Analysis

There are basically two reasons why function fitting is performed on data that comes from measurement tools like SEMs, etc. The first one is the need to handle measurement errors that are inherent in raw data. With a curve fit it is possible to smooth out noise in a way that is supported by basic laws of statistics. The other reason is that a curve fit provides an equation on which the powerful analysis tools of algebra can be used. In the previous sections the functions for each type of data have been introduced. Each of them depends on a set of parameters, called coefficients, that have to be determined for the actual data set.

Given a set of measured points (x_i,y_i) with x_i the measurement position vector (i.e., input parameter values) and y_i the measured value (output) for each measurement *i*, let *F(x)* be the function to be fitted. The most common way to determine the coefficients is to calculate those coefficients which optimize the merit function

$$\chi^2 = \sum_i \left(y_i - F(x_i)\right)^2 \tag{27}$$

Summing the squares of the distances at each data point, χ^2 (chi-squared) measures the agreement of the fitting function and the data. If the coefficients are chosen such that chi-squared is minimized, a function with the best average approximation for each data point is found.

A general algorithm to perform this optimization has been invented by Marquardt and will be used as the foundation for the whole procedure of finding the desired coefficients [15]. Marquardt's method requires a first guess of the coefficients. It iterates them until the merit function chi-squared has reached a local minimum. The speed of convergence is greatly enhanced by using the partial derivatives of the fitting function to point in the direction of the minimum.

In order to get best results Marquardt's algorithm has been extended and a number of parameters have been added which allow adjustment to the individual needs and weaknesses of a data set. The extensions, parameters and their purposes are described below.

A. Analysis ranges

Often, experimental data is collected over a range of inputs wide enough to completely encompass the region of interest. If the exact "center" and desired range for the data is not known ahead of time, one frequently makes the experimental range even larger in order to be sure to capture the region of interest. This usually means that data collected at the extremes of the input ranges are less valid than data centered in the range (due to systematic measurement errors, etc.). Also, data at the extremes of the possible output range (CDs that are almost scummed, for example) may be less accurate (or of less interest) than values near the target. As a consequence, the range of data used for the fitting can have a big effect on the goodness of fit. By excluding data outside the range of interest, better fits can be obtained within this range of interest. It is important to remember that the analysis ranges are also part of the analysis result, since the fitted function is naturally limited to these ranges as well. The exceptional case where the analysis ranges are made larger than the original data ranges is called data extrapolation and will be discussed below.

B. Removal of statistically bad data points

A standard statistical method to handle data with large measurement errors ("data flyers") in a curve fit is to perform a second fit after removing those data points that exceed a certain deviation from the firstly obtained function. In other words, Marquardt's algorithm to optimize χ^2 is used two times: first, it calculates the coefficients as mentioned above, using all data points in the analysis ranges. Next, those data points whose deviation from the fitted function exceed a specified tolerance are removed and the algorithm is used again to calculate the final coefficients. A good choice for the deviation tolerance is usually two times the standard deviation σ from the first fit, where the standard deviation is defined as

$$\sigma = \sqrt{\chi^2 / N-1} \tag{28}$$

and where N = number of data points. However, another multiple of σ or the direct selection of a deviation tolerance can sometimes be the better decision.

C. Data point weighting

Some data sets have a center in which the measured values have more importance than values at the edges of the data range. Focus-Exposure matrices especially are measured around an estimated best focus and best exposure and the data closest to the center of the range is most important. A way to represent this in the curve fit is to assign to each data point an individual weight w_i. By optimizing the weighted chi-square,

$$\chi_w^2 = \sum_i w_i^2 \left(y_i - F(x_i)\right)^2 \tag{29}$$

instead of chi-squared, the obtained function will tend to fit data points with more weight more closely than data points with less weight.

A 2D Gaussian distribution is a useful function for generating the weights for the Focus-Exposure data:

$$w(F,E) = \exp\left(-\frac{(F-F_o)^2}{2\sigma_F^2} - \frac{(E-E_o)^2}{2\sigma_E^2}\right) \tag{30}$$

Although this Gaussian distribution is described by four parameters, these four can be reduced to a single user-adjustable term λ by the following assignments:

$$F_o = \frac{(F_{\max} + F_{\min})}{2} = \text{(middle of the focus analysis range)}$$

$$E_o = \frac{(E_{\max} + E_{\min})}{2} = \text{(middle of the exposure analysis range)}$$

$$\sigma_F = \lambda \frac{(F_{\max} - F_{\min})}{2} = \text{(distance from center point to the focus analysis bounds)}$$

$$\sigma_E = \lambda \frac{(E_{\max} - E_{\min})}{2} = \text{(distance from center point to the exposure analysis bounds)}$$

where the single adjustable parameter λ is called the Gaussian "stretch factor". Its value is responsible for how tight the center of importance is and it is usually set to be about 1.0. Increasing λ results in increasing the area with the most weight, decreasing λ results in giving only the points in the very center more weight. Setting λ to be very large (say, greater than 5) is equivalent to turning off Gaussian weighting. Note that $w(F_o, E_o)$ is always equal to one.

D. Individual coefficient manipulation

Sometimes some of the fitting function coefficients are known before the data has been fitted. For swing curve data, for example, one of the coefficients is the period that depends directly on the refractive index of the resist and the wavelength, which are sometimes both known before the data is collected. It is then very useful to assign these coefficients their known values and fix them (that is, do not allow them to be varied in the process of finding the best fit). The remaining coefficients will then be optimized without changing the fixed coefficients, so that the result matches the additional information about the data.

Another reason to fix coefficients to certain values might be the desire to force the fitting function to have a shape that is less general than the equation allows it to be. A good example is the polynomial in equation (23). By fixing the higher order coefficients to zero, the function becomes a polynomial of a lower degree which can help to fit data with large measurement errors or a small number of data points by allowing less flexibility. Also, by fixing coefficients to zero that belong to odd focus terms, a symmetrical fit can be achieved.

E. Restrictions in the optimization procedure

The optimization algorithm from Marquardt works very well when no restrictions are placed on the values of the coefficients. Unfortunately, this is not always practical. Although the general polynomial of equation (23) is extremely flexible at fitting Focus-Exposure CD data, this same flexibility can lead to non-physical results. In reality, the effect of exposure on CD is monotonic – the CD either always increases or always decreases with increasing exposure. A general polynomial fit has no such restrictions, potentially leading to non-physical best fits. Therefore, a special algorithm has been developed that restricts the optimization procedure in such a way to always provide monotonic exposure behavior.

F. Data extrapolation

When the obtained fitting function is analyzed at a position that exceeds the original data ranges, data extrapolation is automatically achieved. The resulting best fit coefficients are not affected by the use of an extrapolated data range except for the special case where the monotic exposure restriction is applied to focus-exposure CD. However, it is worth mentioning that data extrapolation is not recommendable because the results are usually poor.

VI. Conclusions

Data analysis is an important part of the photolithography engineer's job. As linewidth control becomes more critical and process windows become smaller and smaller, accurate analysis of lithography process data becomes essential. Simple techniques, such as plotting swing curve data and estimating the position of a maximum visually, or simply plotting a Bossung curve to guess best focus, is no longer adequate in most manufacturing environments. Automated, statistically sound techniques for analyzing data, removing bad data points, and extracting relevant lithographic information can dramatically improve one's ability to monitor, characterize, and optimize a process.

The techniques presented here have been incorporated into the software tool ProDATA™. This comprehensive lithographic data analysis tool employs the open curve-fit models described above and can form the basis of a standard methodology for many common semiconductor research, development, and manufacturing tasks.

References

1. C. P. Ausschnitt, “Rapid Optimization of the Lithographic Process Window,” *Optical/Laser Microlithography II, Proc.*, SPIE Vol. 1088 (1989) pp. 115-133.

2. P. G. Drennan, “Convenient Optimization of the Lithographic Process Window,” *OCG Microlithography Seminar Interface ’93* (1993) pp. 229 – 241.

3. R. A. Ferguson, R. M. Martino, and T. A. Brunner, “Data Analysis Methods for Evaluating Lithographic Performance,” *J. Vac. Sci. Technol.*, Vol B 15, No. 6 (Nov/Dec 1997) pp. 2387-2393.

4. F. Hurter and V. C. Driffield, “Photochemical Investigations and a New Method of Determination of the Sensitiveness of Photographic Plates,” *Journal of the Society of the Chemical Industry* (May 31, 1890) pp. 455-469.

5. C. A. Mack, “Lithographic Optimization Using Photoresist Contrast,” *KTI Microlithography Seminar, Proc.*, (1990) pp. 1-12, and *Microelectronics Manufacturing Technology*, Vol. 14, No. 1 (Jan. 1991) pp. 36-42.

6. ASTM Standard F1059-87, American Society for Testing and Materials (Philadelphia, PA: 1987).

7. D. H. Ziger and C. A. Mack, “Generalized Approach toward Modeling Resist Performance,” *AICHE Journal*, Vol. 37, No. 12 (Dec 1991) pp. 1863-1874.

8. C. A. Mack, “Development of Positive Photoresist,” *Jour. Electrochemical Society*, Vol. 134, No. 1 (Jan. 1987) pp. 148-152.

9. C. A. Mack, “Absorption and Exposure in Positive Photoresist,” *Applied Optics*, Vol. 27, No. 23 (1 Dec. 1988) pp. 4913-4919.

10. C. A. Mack, “Standing Waves in Photoresist,” *Microlithography World*, Vol. 3, No. 2 (Spring 1994) pp. 22-24.

11. C. A. Mack, “Swing Curves,” *Microlithography World*, Vol. 3, No. 3 (Summer 1994) pp. 23-25.

12. J. W. Bossung, “Projection Printing Characterization,” *Developments in Semiconductor Microlithography II, Proc.*, SPIE Vol. 100, pp. 80-84 (1977).

13. C. A. Mack, “Understanding Focus Effects in Submicron Optical Lithography: a Review,” *Optical Engineering*, Vol. 32, No. 10, pp. 2350-2362 (Oct., 1993).

14. E. W. Charrier and C. A. Mack, “Yield Modeling and Enhancement for Optical Lithography,” *Optical/Laser Microlithography VIII, Proc.*, SPIE Vol. 2440, pp. 435-447 (1995).

15. D. W. Marquardt, “An Algorithm for Least-Squares Estimation of Non-linear Parameters,” *J. Soc. Indust. Appl. Math.*, Vol. 11, No. 2 (June, 1963) pp. 431-441.