Improved Model for Focus-Exposure Data Analysis

Chris A. Mack and Jeffrey D. Byers

KLA-Tencor, FINLE Division
8834 N. Capital of Texas Highway, Suite 301, Austin, TX 78759 USA
e-mail: chris.a.mack@kla-tencor.com

Abstract

The paper introduces an improved, physics-based function for fitting lithographic data from focus-exposure matrices. Unlike simple polynomial functions, the coefficients of this equation offer physical insight into the meaning and nature of the data. Derivation of this equation from first principles of the physics of lithographic imaging is presented. Examples based on typical experimental data are shown and the advantages of using a physics-based fitting function is described based on improved fitting and noise filtering.

Keywords: Focus-exposure matrix, process window, data analysis, ProDATA

I. Introduction

Systematic analysis of focus-exposure matrix data is vital to the accurate determination of process windows and the calculation of depth of focus and best focus [1-3]. This analysis is generally accomplished by first fitting the data to a mathematical function, then using this function for process window determination. The advantage of this approach is that the goodness of fit can be used as an objective means of data flier removal, and the natural "smoothness" of the fitting function can reduce the impact of experimental noise in the data on process window determination. Improper selection of the fitting function, however, can lead to other problems. A function with two few parameters could eliminate real and significant patterns in the data. A function with two many parameters can produce artifacts that do not actually exist in the data. Choosing the correct function, with the correct number and type of fitting coefficients, is critical to proper process window determination.

The paper introduces an improved, physics-based function for fitting lithographic data from focus-exposure matrices. Unlike simpler polynomial functions, the coefficients of this equation offer physical insight into the meaning and nature of the data. Derivation of this equation from first principles of the physics of lithographic imaging is presented. Numerous examples based on typical and unusual experimental data will be shown and the advantages of using a physics-based fitting function is described.

II. Polynomial Focus-Exposure Matrix Data Analysis

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 1 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 [4]. 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

$$CD = \sum_{i=0}^{3}\sum_{j=0}^{4} a_{ij} E^i F^j \qquad (1)$$

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.

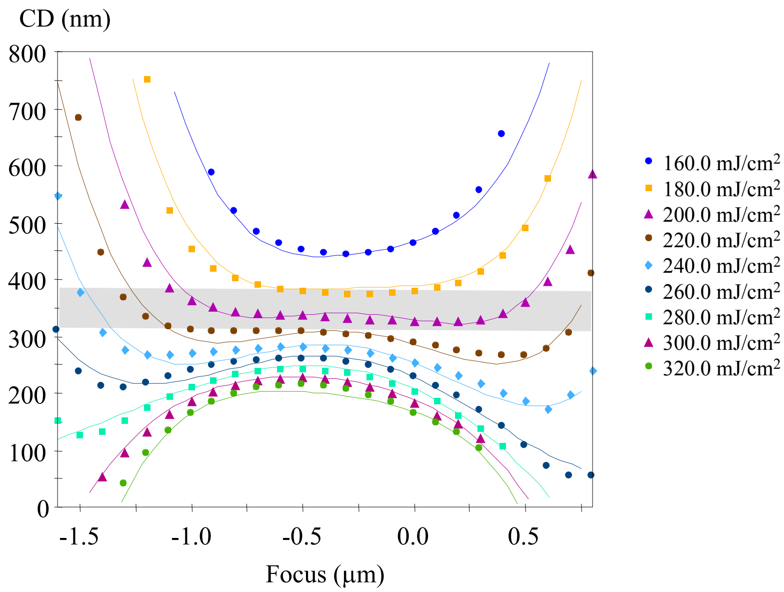


Figure 1. Example of the effect of focus and exposure on the resulting resist linewidth (symbols) and the best fit of this data (lines) to equation (1).

III. Curve Fitting and Statistical Analysis

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 that optimize the merit function

$$\chi^2 = \sum_i \left(y_i - F(x_i)\right)^2 \qquad (2)$$

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 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, an 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

$$\sigma = \sqrt{\chi^2 / N-1} \tag{3}$$

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

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,

$$\chi_w^2 = \sum_i w_i^2 \left(y_i - F(x_i)\right)^2 \tag{4}$$

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

One approach to data weighting is weight each data point in inverse proportion to the uncertainty in the data. If repeat measurements are made, either on a single experimental observation or on repeat experiments, the statistics of the measurements will produce a standard deviation which can be used as an estimate of the uncertainty of the data point. By weighting each point as one over the standard deviation of the measurement, the most certain points will have the greatest influence on the fit. In practice, the out of focus features, with their poor resist profiles and higher sensitivity to process variations, will have greater uncertainty and thus will, in general, be weighted less.

IV. Improved Physics-Based Fitting Function

Using a simple polynomial to fit experimental focus-exposure CD data can have certain problems. Using the polynomial function with two few parameters could eliminate real and significant patterns in the data. Using the function with two many parameters can produce artifacts that do not actually exist in the data. It is often difficult, over a wide range of data sets, to determine the best number of coefficients to use in the fit. A physically based fitting function can aid in this decision, with the added benefit of physical meaning for the coefficients. Additionally, a well designed physically based fitting function should allow the best fit with the fewest number of adjustable parameters, thus increasing the confidence in data flier removal and in preserving the integrity of true data patterns.

Consider first the aerial image of a simple pattern of lines and spaces of pitch p. In one dimension (x), the aerial image can always be expressed as a Fourier series.

$$I(x) = \beta_0 + \beta_1 \cos(2\pi x / p) + \beta_2 \cos(4\pi x / p) + ... \tag{5}$$

As the simplest example, consider equal lines and spaces of width w (= $p/2$). Note that the derivation that follows is conceptually simpler for the case of equal lines and spaces, but the results are not limited to this case. It will be convenient to replace the x position with a coordinate Δw that is the deviation of x from the nominal line edge.

$$x = \frac{w}{2} + \frac{\Delta w}{2} \tag{6}$$

Writing equation (5) in terms of this new coordinate gives

$$I(x) = \beta_0 - \beta_1 \sin\left(\frac{\pi}{2}\frac{\Delta w}{w}\right) - \beta_2 \cos\left(\pi \frac{\Delta w}{w}\right) - \ldots \tag{7}$$

Eventually we will use this equation to predict the behavior of feature size with focus and exposure. Since the features of interest will be near the nominal size, we will be most interested in equation (7) for the case of small $\Delta w/w$, with values between –0.2 and 0.2 of greatest interest. Thus, it will be reasonable to expand the sine and cosine terms of this equation in a Taylor series, keeping only the first few terms.

$$I(x) \approx \gamma_0 - \gamma_1\left(\frac{\Delta w}{w}\right) + \gamma_2\left(\frac{\Delta w}{w}\right)^2 + \gamma_3\left(\frac{\Delta w}{w}\right)^3 - \ldots \tag{8}$$

where the γ coefficients are just linear combinations of the β terms (for example, $\gamma_0 = \beta_0 - \beta_2$, etc.). Although the exact values of the γ and β terms are a function of the pitch, wavelength, numerical aperture, and partial coherence, for the simple coherent illumination case and equal lines and spaces $\gamma_0 = 0.25$, $\gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 0.4112$, etc.

Given an aerial image one can estimate the change in critical dimension (CD) as a function of exposure dose, E. Using the simple approximation of a thin, infinite contrast photoresist, the photoresist will be removed whenever the dose exceeds some threshold dose, E_{th}. In other words,

$$EI(\Delta w = CD - w) = E_{th} \tag{9}$$

Keeping only through the second order terms in equation (8), one can solve for the resulting CD as a function of dose.

$$\frac{CD - w}{w} = \frac{\Delta w}{w} \approx \frac{\gamma_1}{2\gamma_2}\left(1 - \sqrt{1 - \frac{4\gamma_2}{\gamma_1^2}\left(\gamma_0 - \frac{E_{th}}{E}\right)}\right) \tag{10}$$

Again, for small $\Delta w/w$ the argument of the square root must necessarily take the form of one minus a small number. Noting that the dose to size, E_s, must be equal to E_{th}/γ_0 and taking the Taylor expansion of the square root,

$$\frac{\Delta w}{w} = \frac{CD - w}{w} \approx \frac{\gamma_0}{\gamma_1}\left(1 - \frac{E_s}{E}\right) + \frac{\gamma_2 \gamma_0^{\,2}}{\gamma_1^{\,3}}\left(1 - \frac{E_s}{E}\right)^2 + \ldots \tag{11}$$

Putting equation (11) into a more general form and keeping only the first N terms in the series,

$$\frac{CD - w}{w} = \sum_{n=1}^{N} c_n \left(1 - \frac{E_s}{E}\right)^n \tag{12}$$

In general, keeping three terms or less in this series gives very good fits to all data sets that we have observed. For many data sets, keeping only the first term provides adequate results.

As an interesting aside, equation (10) simplifies to the following expression for the case of coherent illumination of small equal lines and spaces:

$$\frac{CD - w}{w} \approx \frac{1}{2}\left(1 - \sqrt{\frac{E_s}{E}}\right) \tag{13}$$

Before continuing with the derivation and adding a focus dependence to this equation, it is instructive to consider the lithographic significance of the coefficients c_n in equation (12). The coefficient c_1 represents the slope, on a log-log scale, of the CD versus dose curve (and is thus the inverse of exposure latitude).

$$c_1 = \frac{\gamma_0}{\gamma_1} = \left.\frac{\partial \ln CD}{\partial \ln E}\right|_{E=E_S} \propto \frac{2}{NILS} \tag{14}$$

where NILS is the normalized image log-slope (the proportionality to NILS is in fact an equality in the limit of an infinite contrast photoresist). The second order coefficient c_2 represents the curvature of the log-CD versus log-dose curve, which is the change in exposure latitude with exposure dose.

By noting the relationship between c_1 and NILS, it becomes possible to incorporate the impact of focus errors on CD. It is well known that NILS falls off with increasing defocus. For a small line/space pattern with coherent illumination, the behavior of NILS with defocus distance δ is just

$$NILS = NILS_{in\ focus} \cos(\pi \delta \lambda / p^2) \tag{15}$$

Keeping with our theme, we will again expand the cosine term with a Taylor series. Thus, the first coefficient of equation (12) will become a function of focus as

$$c_1 = \left.\frac{\partial \ln CD}{\partial \ln E}\right|_{in\ focus} + \alpha \delta^2 + \ldots \tag{16}$$

Although in the ideal case equation (16) has only even powers of defocus distance, real lithographic results do exhibit some asymmetry with defocus. Thus, a generalization of equation (16) would become

$$c_n = \sum_{m=0}^{M} b_{nm} (F - F^*)^m \tag{17}$$

where F is the focal position, F^* is best focus and the defocus distance is F-F^*. In general, M = 4 is sufficient to describe the vast majority of data sets and often the odd terms are very small or negligible.

Combining equations (12) and (17) and simplifying,

$$CD = \sum_{m=0}^{M} \sum_{n=0}^{N} a_{nm} \left(1 - \frac{E_s}{E}\right)^n F^m \tag{18}$$

Equation (18) represents a physically based fitting function that has improved fitting performance over the original polynomial formulation of equation (1). As others have noted [5,6], a CD variation with one over exposure dose is more physically accurate than assuming that CD is proportional to powers of dose.

V. Applying the Improved Fitting Function

One of the goals of using the fitting function described above rather than the original polynomial expression is to provide better fits of experimental data with fewer terms (and thus fewer adjustable fitting coefficients). Doing so should lead to improved tolerance to statistical noise in the data and better data flyer removal decisions. In order to test out these attributes of a data fitting function an ideal "noise free" data set was generated using simulation. A typical 248nm chemically amplified resist process was used to create 130nm dense line/space focus-exposure matrix data sets.

Figure 2 shows example fits of two data sets using the simple polynomial and the new physically based fitting function. The two sets are identical except that the second data set used a wider range of focus. In each case the number of adjustable parameters in the two fitting functions was kept the same. For the limited focus range case, the original polynomial with six terms fit the data with a one sigma goodness of fit of 3.49nm. The physically based fitting function showed a goodness of fit of 1.47nm. For the extended focus range case, using a greater number of terms in each expression to capture the more interesting focus behavior, the physically based function again resulted in a much smaller goodness of fit, 1.90nm versus 4.26nm for the original polynomial.

To test the robustness of each expression with respect to noise in the data, random noise of various amounts was added to the first data set and fits using both fitting expressions were repeated. To make the "noise" as realistic as possible, a metrology noise was added to the CD values itself, but noise was also added to the focus and exposure values in the data set. Since a random number generator was used to generate the added noise, several noisy data sets were generated for each nominal magnitude of added noise and then the actual RMS noise amount was measured from the result. Figure 3 shows example fits when 4.1nm RMS of random Gaussian noise was added to the data set. The goodness of fits were 4.53nm and 4.24nm for the original and improved fitting functions, respectively. Note that the advantage that the improved function enjoyed in terms of goodness of fit seems to be washed away by the noise in the data. The effect of noise on the goodness of fit is explored in more detail in Figure 4, where the RMS magnitude of added noise is varied and the goodness of fit of each expression is plotted. One can see that the six term physically based function out performs the six term simple polynomial at low noise levels, and in fact is comparable to a twelve term polynomial fit. However, at high noise levels all functions give approximately the same goodness of fit.

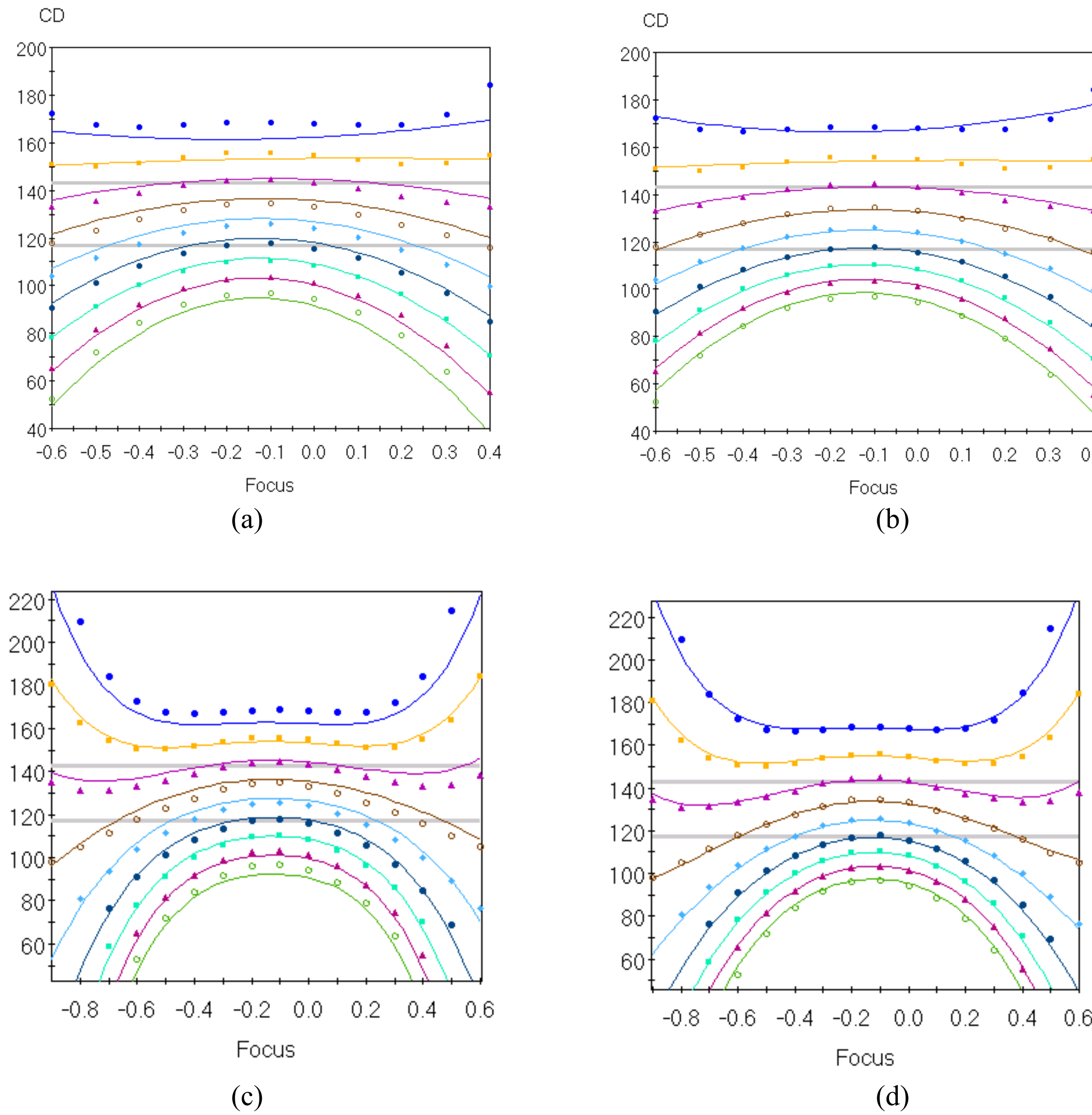
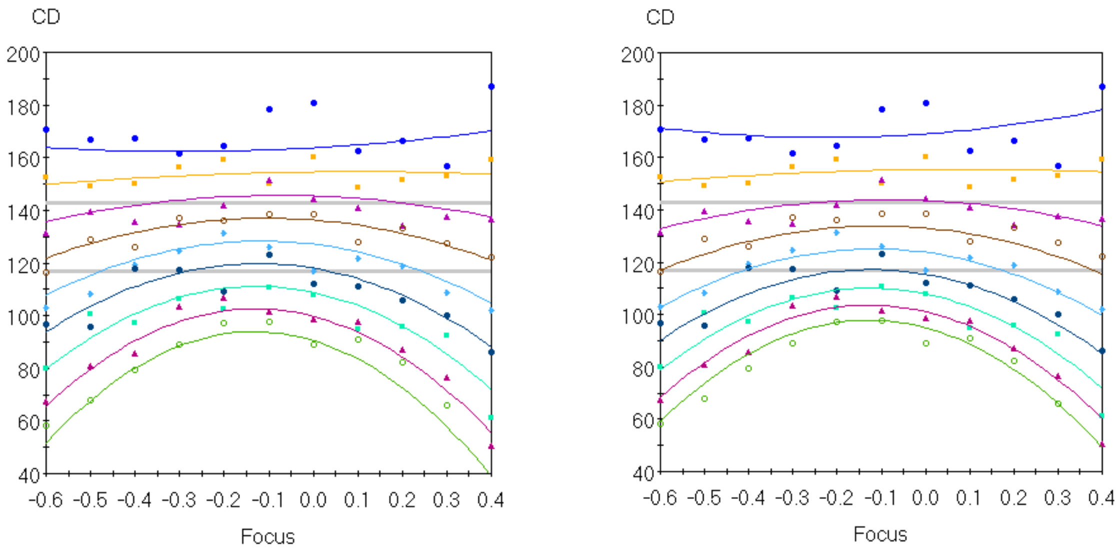


Figure 2. Comparison of the original (a and c) to the new (b and d) fitting function using the same number of terms in each expression and “noise free” data generated by simulation.



(a) (b)

Figure 3. Comparison of the original polynomial (a) to the physically based fitting function (b) to data with 4.1nm RMS of added noise.

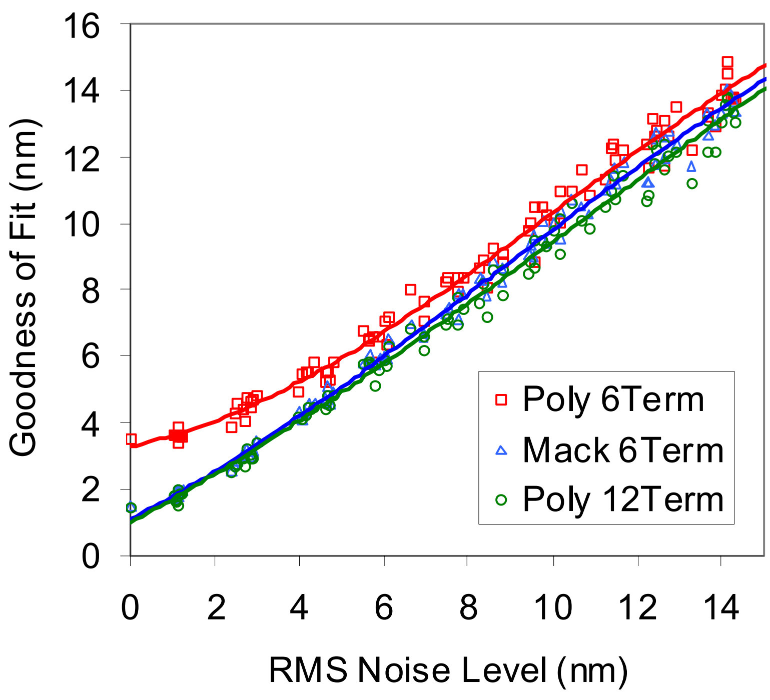


Figure 4. Comparison of goodness of fits for the various fitting functions in the presence of noise added to the data.

But is the goodness of fit to noisy data the best metric of the appropriateness of a function to the task of data fitting? A goal of using a fitting function to describe experimental data is to filter out noise and extract the true, core behavior present in the data. In general one does not know the true behavior of experimental data in the absence of noise. In our case, however, the data has been generated by adding a set amount of noise to an ideal noise free data set. Thus, a more appropriate metric for how well the fitting function filters out random noise in our experiment would be to use an RMS model error to judge the result:

$$\text{RMS Model Error} = \sqrt{\frac{\sum(\text{Model} - \text{Noise Free Data})^2}{\text{N}}} \tag{18}$$

Note that the model is first fit to the noisy data in the standard way, by minimizing the goodness of fit to the noisy data. Then, the effectiveness of the model is measured using the RMS model error of equation (18).

Figure 5 shows the results. As can be seen the physically based model does a much better job of filtering out noise and keeping the RMS model error low in the presence of large amounts of noise. In fact, the six term physically based function does a slightly better job of preserving the original noise free behavior than even the twelve term polynomial. Adding more polynomial terms results in the fitting of noise, giving a goodness of fit that is below the amount of added RMS noise. As a result, the higher term polynomial may in fact have a worse RMS model error than the six term polynomial when a high noise level is present.

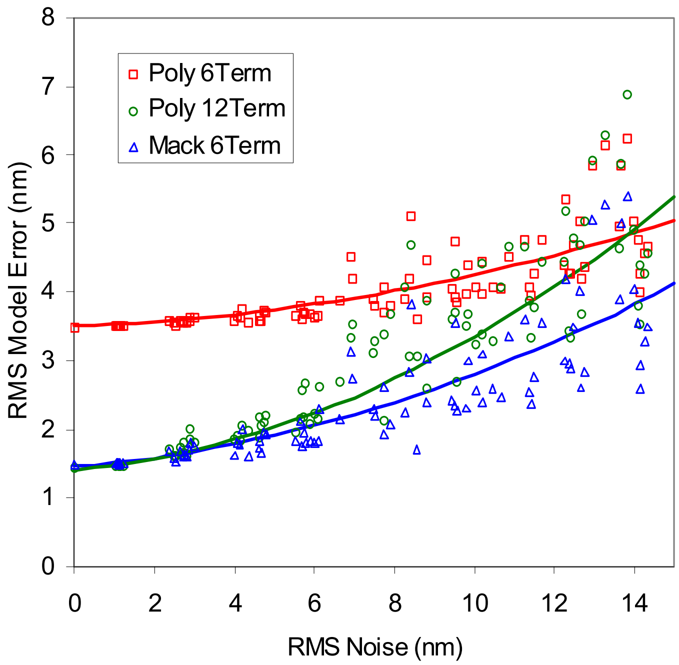


Figure 5. Behavior of the various fitting functions in the presence of added noise using the RMS model error as the metric to judge the effectiveness of each model at filtering noise.

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. 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.

This paper introduces an improved physics-based function for fitting lithographic data from focus-exposure matrices. Unlike simple polynomial functions, the coefficients of this equation offer physical insight into the meaning and nature of the data, based on its derivation from first principles of the physics of lithographic imaging. The advantages of using a physics-based fitting function was shown based on improved fitting and noise filtering. The improved physics based fitting function presented here has been incorporated into the software tool ProDATA™.

References

1. C. A. Mack, D. A. Legband, S. Jug, "Data Analysis for Photolithography" *Micro- and Nano-Engineering 98, Proc.*, (1998) pp. 65-68.
2. C. A. Mack, S. Jug, D. A. Legband, "Data Analysis for Photolithography," *Metrology, Inspection, and Process Control for Microlithography XIII, Proc.*, SPIE Vol. 3677 (1999) pp. 415-434.
3. S. Jug, R. Huang, J. Byers, C. A. Mack, "Automatic Calibration of Lithography Simulation Parameters," *Lithography for Semiconductor Manufacturing, Proc.*, SPIE Vol. 4404 (2001).
4. J. W. Bossung, "Projection Printing Characterization," *Developments in Semiconductor Microlithography II, Proc*., SPIE Vol. 100, pp. 80-84 (1977).
5. C. P. Ausschnitt, "Rapid Optimization of the Lithographic Process Window," *Optical/Laser Microlithography II, Proc.*, SPIE Vol. 1088 (1989) pp. 115-133.
6. D. Fuard, M. Besacier, P. Schiavone, "Assessment of Different Simplified Resist Models," *Optical Microlithography XV, Proc.*, SPIE Vol. 4691 (2002).