This paper was published in the proceedings of
Design, Process Integration, and Characterization for Microelectronics
SPIE Vol. 4692, pp. 454-464.

Characterizing the Process Window of a Double Exposure Dark Field Alternating Phase Shift Mask

Chris A. Mack

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

After reviewing the normal approach to process window characterization and analysis used for a standard single exposure process, the applicability of that approach to a double exposure process is investigated. By properly designing the binary trim mask for a double exposure dark field alternating phase shift mask process, the influence of the trim exposure step on the final gate CD can be minimized. Flare during the trim exposure, however, is found to cause an undesired coupling of the two exposure steps. A method of accounting for this coupling effect for the gate CD process window is given.

Keywords: Phase Shift Mask, Process Window, Lithography Simulation, PROLITH

I. Introduction

While alternating phase shift masks offer the ultimate resolution with large process windows for certain types of features, phase conflicts and phase termination problems make their design for arbitrary circuit patterns exceedingly difficult. One simple, though costly, solution is to use a double exposure: one mask which has optimum phase assignments without regard to the termination problem, and a second “trim” mask to fix the termination problems by exposing them away. The flexibility in mask layout for such a double exposure scheme offers several types of optimizations. For example, the phase mask pattern can be made as a dark field mask, thus getting the advantages of reduced reticle defect sensitivity and simple layout rules. For logic gate pattern masks, this phase shift region can be made to coincide with the intersection of the gate and the active area, thus providing a simple design paradigm that naturally leads to a minimum number of phase conflicts. As a result, double exposure dark field alternating phase shift masks have received considerable attention, despite the costs associated with the reduced litho tool throughput.

Since the major benefit of phase shift masks is the enlargement of the focus-exposure process window, an interesting question arises: How does one characterize the process window of a double exposure process? After all, there are now two focal positions and two exposure doses. Is a four-dimensional process window required? Are the sources of focus and exposure dose errors common between the two exposure passes, or are they independent? The answers to these questions will lead to a recommended procedure for the characterization and understanding of the focus-exposure process window of a double exposure process. Using PROLITH simulations the impact of the trim mask exposure and focus settings on the final gate CD will be evaluated.

II. The Focus Exposure Process Window – A Review

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 focus-exposure data for a standard, single exposure process.

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, develop time, and reticle errors. Of these secondary process errors, one of the most important is exposure.

Since the effects of focus are 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 [1].

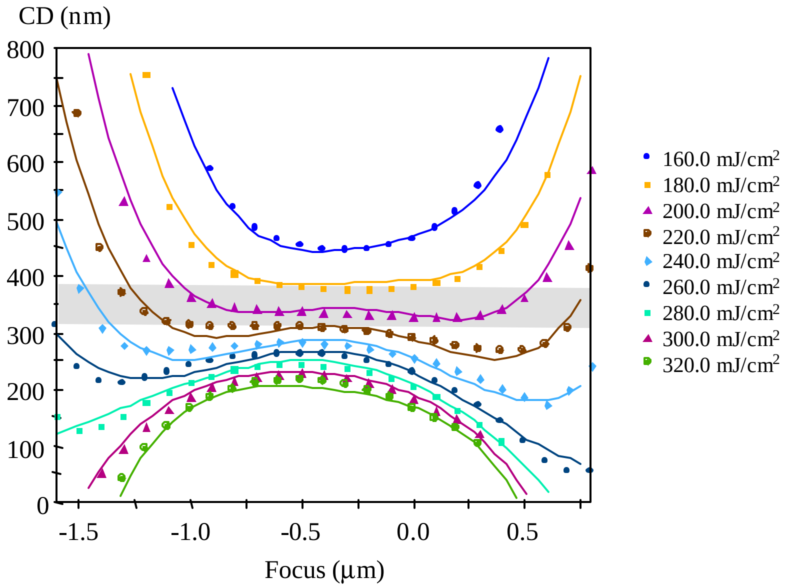


Figure 1. Example of the effect of focus and exposure on the resulting resist linewidth.

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 2).

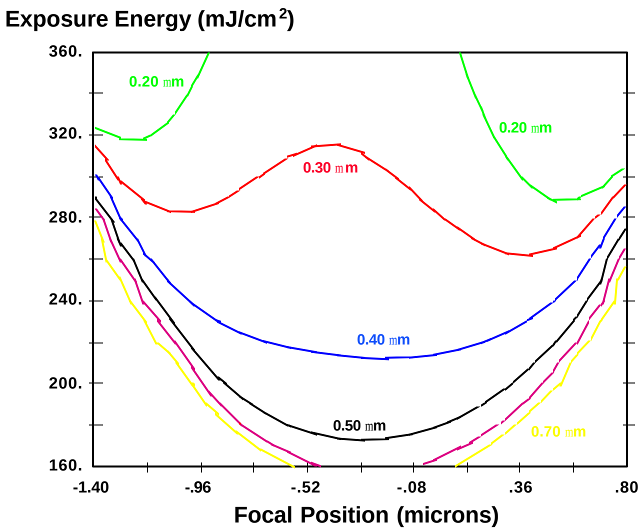
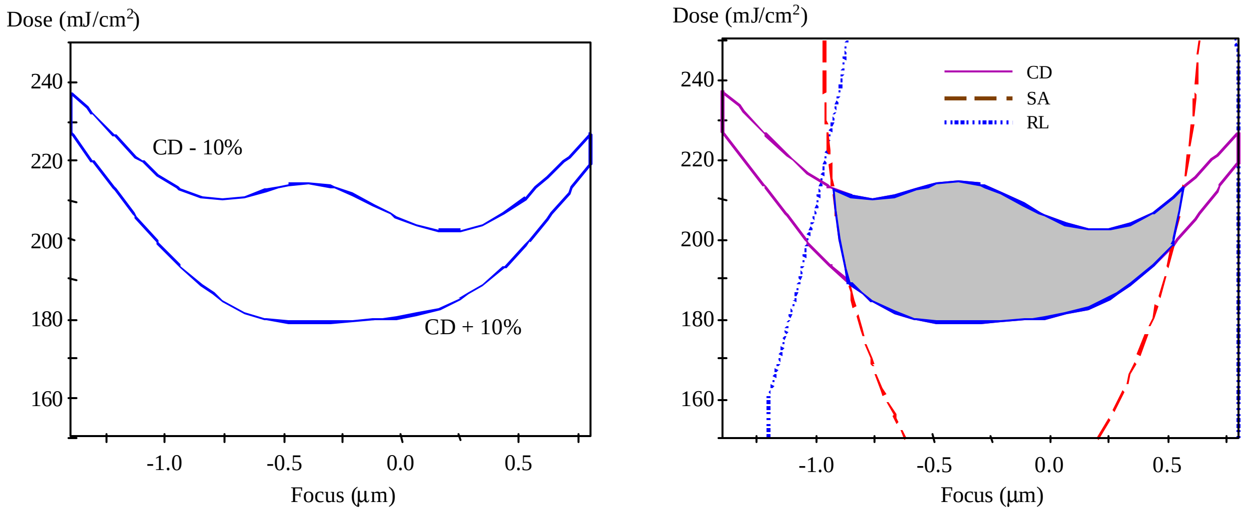


Figure 2. Displaying the data from a focus-exposure matrix in an alternate form, contours of constant CD versus focus and exposure.

The contour plot form of data visualization is especially useful for establishing the limits of exposure and focus that allow the final 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 3a). Because of the nature of a contour plot, other variables can also be plotted on the same graph. Figure 3b 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* [2]. 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 [3], but a simpler analysis can give useful insight and can be used to derive a number for depth of focus.



(a) (b)

Figure 3. 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 rectangles that will fit inside the processes window. Figure 4 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 4.

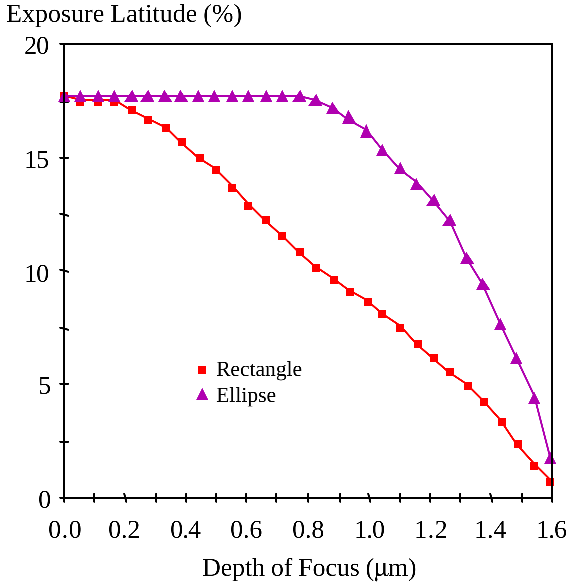


Figure 4. The process window of Figure 3 is analyzed by fitting all the maximum rectangles and all the maximum ellipses, then plotting their height (exposure latitude) versus their width (depth of focus).

The exposure latitude versus DOF curves of Figure 4 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 4, 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) = 0.85 µm

DOF (ellipse) = 1.35 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 fits 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 across field. Figure 5 shows a simple example of two overlapping CD process windows for dense and isolated lines.

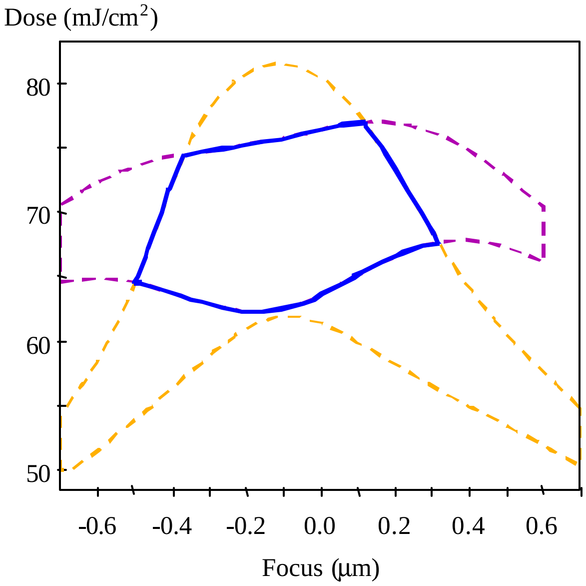


Figure 5. An example of overlapping process windows for dense and isolated lines.

Two important points should be made concerning the above discussion. First, the nature of the Bossung plot and the resulting process window recognizes that the variables focus and exposure for a standard single exposure process are *coupled* in their effect on the lithographic process. The analysis of the size of the process window by fitting ellipses or rectangles inside assumes that the sources of focus and exposures errors are completely *independent* of each other. In other words, going out of focus does nothing to increase or decrease the magnitude of exposure errors. Likewise, and error in exposure has no impact on the sources of focus errors. An analysis of the typical sources of focus and exposure errors confirms that this is an accurate assumption.

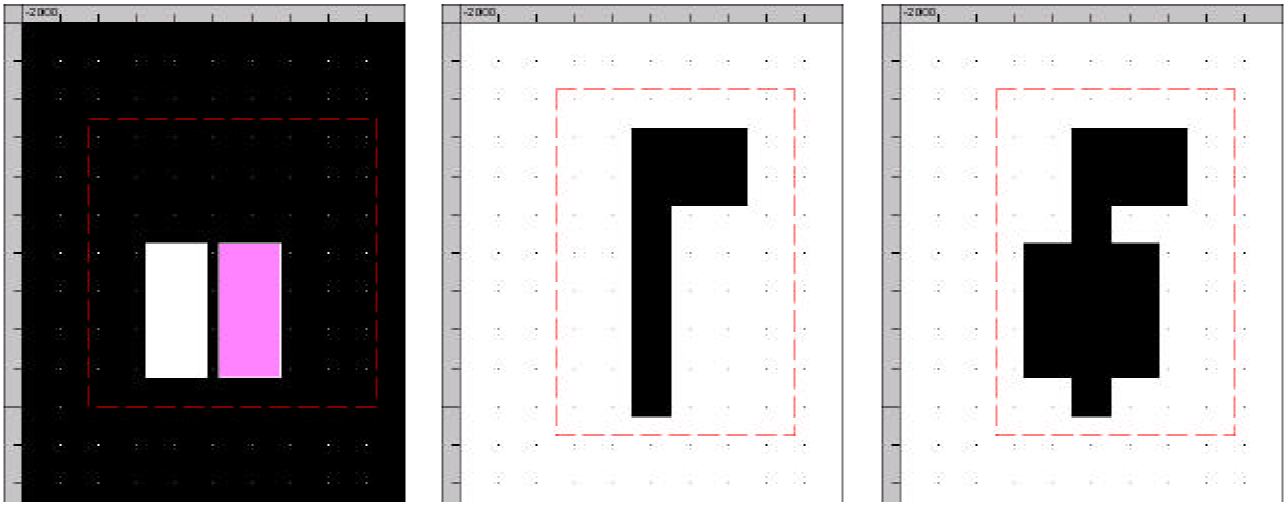
II. Double Exposure Process Window Effects

How can the process window analysis method described above be applied to the more complicated case of a double exposure process? For a double exposure process there are two exposure values and two focus values. The simple-minded approach would be to expand the process space to four dimensions. Although cumbersome and hard to visualize, a four input process window is technically feasible if it were absolutely required. Analyzing the size of the window, however, would present additional problems. The assumption, used above for the standard process window, that errors in each input parameter are completely independent is certainly not true for the case of a double exposure 4D process window. Many of the focus errors, such as wafer non-flatness or field curvature would be coupled between the two exposure passes. Defining the DOF for each pass would be a complicated matter.

If, however, the two sets of focus and exposure settings were independent in their influence on the critical dimension of the gate being produced, the answer would be easy. Two separate process windows could be generated and analyzed independently, giving a DOF for each exposure step. As a practical matter, complete independence is not required. If the less critical exposure step (the binary mask exposure, for example) does not influence the process window of the critical exposure step (the PSM exposure) over the range of typical exposure and focus error values expected for the non-critical step then the two sets of inputs could be assumed independent of each other.

Is such an assumption valid? Although the answer will depend on the specifics of the mask layout and process settings, two typical dark field phase shifting double exposure masks with two different processes will be examined through simulation to determine the answer for these cases. For all cases the dark field phase shifted mask has a semi-isolated phase region 1400nm wide by 1000nm tall (wafer dimensions) with a 120nm chrome line down the center separating the two phase regions (Figure 6). Two separate trim masks were used. The first (called the “small chrome” trim mask) covered the active area with a chrome line 400nm wide (to protect the phase edge produced gate), while the second mask (“large chrome”) covered the entire active area with chrome.

The exposure tool was a 248nm, 0.7 NA stepper with a partial coherence of 0.3 used for the phase exposure and a value of 0.6 used for the trim mask. The resist simulated was 335nm of Shipley UV6 on 62nm on AR2 on silicon. The target CD used was the natural linewidth of the 0-180° phase edge (= $0.25\lambda/NA$), which is 90nm for this system. A ±10% CD specification and a 10% exposure latitude specification were used for the process window calculations. Figure 7 shows a typical result of the simulation.



(a) (b) (c)

Figure 6. Diagrams of (a) the phase mask and the two trim masks ((b) the small chrome mask, and (c) the large chrome mask) used for the double exposure simulations. Black represents chrome.

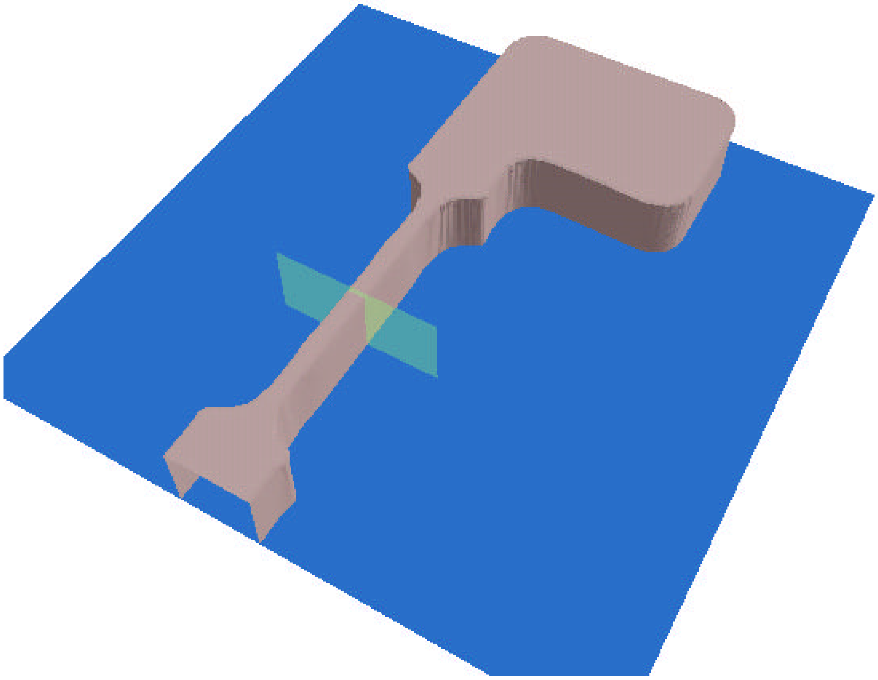
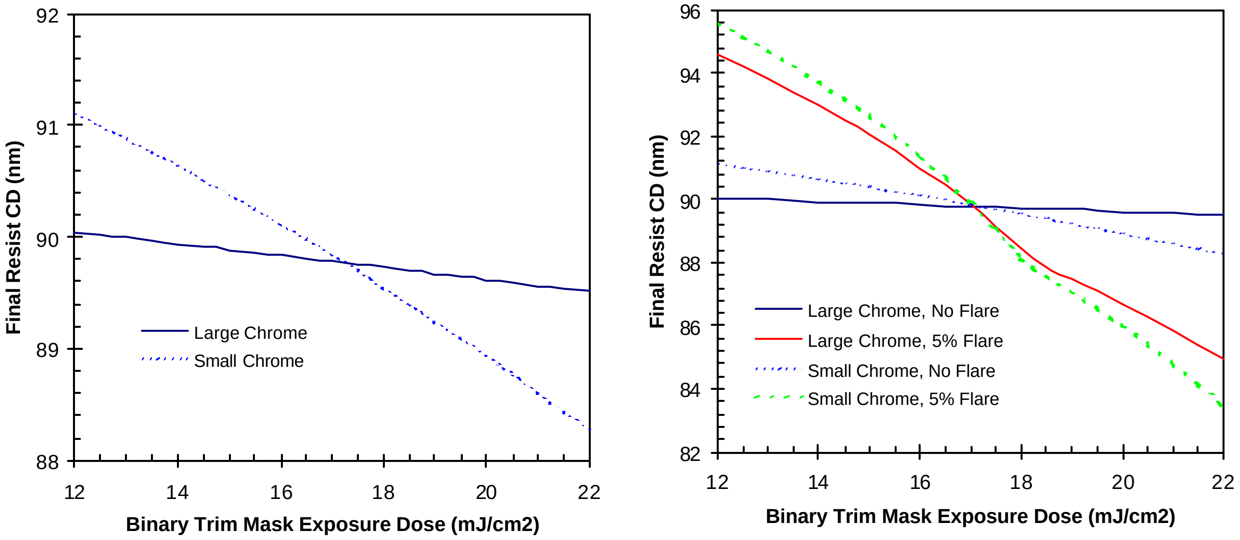


Figure 7. PROLITH simulation of the resist pattern resulting from a double exposure dark field phase shifting mask process.

As an initial way of determining the influence of the trim mask exposure on the gate CD, a plot of gate CD versus trim mask exposure dose was generated. Figure 8a shows the results for the two different trim masks. For the small chrome mask, a 10% range of dose variations for the trim mask exposure will result in a very small 0.5nm CD change. For the large chrome mask, the trim mask exposure had a completely insignificant impact on the final gate CD (a 10% dose error producing only a 0.1nm CD change).

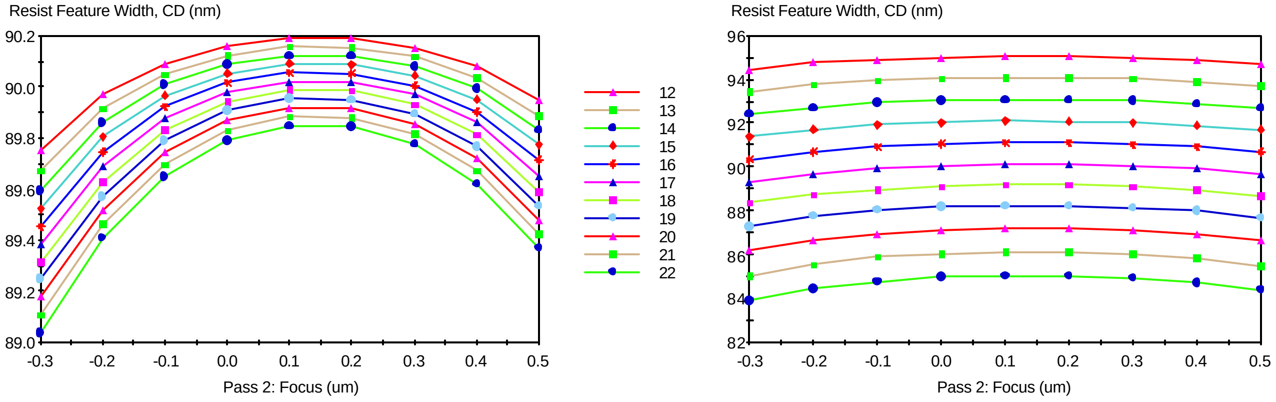
The results change considerably if flare is added into the simulations. Flare, unwanted scattering within the optical system, will be essentially zero for the dark field exposure but may be significant for the bright field trim exposure. The above simulations assume zero flare. If 5% flare is added (a reasonably typical value), this flare will add an extra dose dependence to the gate CD since the amount of flare that actually exposes the gate area will depend on the trim mask dose. Figure 8b shows how the trim mask gate CD exposure latitude worsens in the presence of flare. The impact of a 10% trim mask dose error will be 2.2 and 2.8 nm for the large and small chrome masks, respectively, amounts that are small but not insignificant for a 90nm gate width. Obviously flare has a significant influence on the trim dose dependence.

Finally, the affects of focus for the trim mask depend greatly on the design of the trim mask. For the small chrome mask, defocus resulted in a significant coupling of the trim exposure step to the final gate CD even in the absence of flare. For the large chrome mask, a focus error for the trim mask only slightly increased the exposure coupling to the gate CD (Figure 9).



(a) (b)

Figure 8. The impact of trim mask exposure dose on the final gate CD, a) with no flare, and b) compared to the case of 5% flare.



(a) (b)

Figure 9. For the large chrome mask, defocus of the trim mask has only a slight impact on the coupling of the trim exposure to the gate CD: (a) no flare, (b) 5% flare.

III. New Double Exposure Process Window Analysis Method

The simulations described above indicate the trim mask exposure energy affects the final gate CD due to the coupling effect of flare. Also, the impact of trim mask dose on the final CD appears to be quite linear. As a result, a fairly simple approach to evaluating the impact of trim mask exposure errors on the process window of the phase mask exposure can be used by converting trim mask dose errors into effective phase mask dose errors. That is, an error in dose during the trim mask step has roughly the same effect as adding a scaled dose error during the phase mask exposure step. The following outlines a procedure for including the effects of trim mask dose errors when analyzing the phase mask process window.

First, measure the process window for the phase mask exposure step in the normal way, varying focus and exposure for the phase mask exposure step but keeping the trim mask process constant. Second, measure the "trim mask exposure latitude" of the gate CD by varying the trim mask exposure and keeping all other variables constant at their nominal values. (For a worst case estimate, this trim mask exposure latitude can be measured with the trim mask defocused by an amount equivalent to the expected maximum focus errors in the process.) Next, compare the slope of this trim mask exposure latitude curve to the slope of the phase mask exposure latitude curve (gate CD versus phase mask exposure dose at best focus). The ratio of these slopes will form a dose scaling factor for accounting for trim mask dose errors:

$$DoseScale = \frac{trimmask\ \Delta CD / \%dose\ change}{phasemask\ \Delta CD / \%dose\ change} \tag{1}$$

Any dose error that occurs during the trim exposure step, when multiplied by this dose scale, will behave like a dose error of that magnitude during the phase mask exposure step. Thus, the process window itself is not affected by the double exposure process, but rather the evaluation of the built in dose errors used to determine the acceptability of the process window is affected. For example, using the large chrome trim mask and the process above and assuming 5% flare, the resulting CD versus trim mask exposure plot has a slope of 0.22nm/% dose change. The same plot for the phase mask exposure has a slope of 0.85nm/% dose. The resulting dose scale is 0.26. Thus, any dose errors during the trim mask step can be converted into effective phase mask dose errors by multiplying by 0.26. Essentially, trim mask dose errors for this process are four times less important than phase mask exposure dose errors.

As has been noted above, the magnitude of the flare determines the coupling of the trim mask dose error to the gate CD. Lower values of the flare will result in lower dose scale factors. In fact, the dose scale will be directly proportional to the amount of flare present during the trim mask exposure. Thus, any efforts to reduce flare during this exposure step will result directly in reduced effective dose errors and improved CD control.

IV. Conclusions

A convenient and useful method for evaluating the process window of a typical double exposure dark field phase shifting mask process has been demonstrated. Using simulation to provide insight into the mechanisms, a well designed trim mask, used in the absence of flare during the exposure step, provides nearly complete independence of focus and exposure errors for the two exposure steps. When flare is present, however, the flare provides a coupling of dose errors from the trim mask step to the phase mask imaging step. Focus errors during the trim mask step, however, seem not to have significant influence on the phase mask exposure or the final gate CD for a well designed trim mask.

A simple approach for quantitatively assessing the impact of trim mask dose errors on the process window and resulting analysis of the phase mask exposure step has been proposed. By using the ratio of the slopes of the gate CD versus exposure dose of the two exposure steps, a dose scale factor is produced. Errors in dose that occur during the trim mask step are multiplied by the dose scale factor to create effective phase mask dose errors. These errors can be used in the evaluation of the process window by helping to determine the minimum acceptable exposure latitude used for depth of focus determination.

Finally, the dose scale factor is (roughly) linearly proportional to the flare in the exposure tool. Efforts to reduce flare will directly impact the effective dose errors and the resulting gate CD control for these double exposure processes. As an aside, a double exposure process might be quite useful as a means of measuring flare in a projection imaging tool by taking advantage of the change in exposure latitude due to the coupling effects of flare.

References

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

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

3. 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).