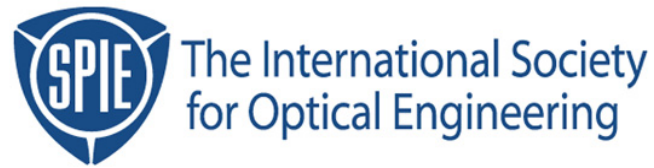


Copyright 2003 by the Society of Photo-Optical Instrumentation Engineers.



This paper was published in the proceedings of
Optical Microlithography XVI, SPIE Vol. 5040, pp. 151-161.
It is made available as an electronic reprint with permission of SPIE.

One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Measuring and Modeling Flare in Optical Lithography

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

Flare, unwanted scattered light arriving at the wafer, is caused by anything that forces the light to travel in a “non-ray trace” direction. The amount of flare experienced by any given feature is a function of both the local environment around that feature (short range flare) and the total amount of energy going through the lens (long range flare). This paper discusses the various sources of flare and reviews the many techniques used to measure flare in lithographic imaging tools. Flare will be described by a new “DC” or low frequency model based on a scattering mechanism that properly accounts for conservation of energy and which improves upon existing DC flare models.

Keywords: flare, scattered light, lithography modeling, PROLITH

I. Introduction – What is Flare?

The goal of an imaging lens is to collect diffracted light, spreading out away from an object, and focus that light down to an image plane, creating an image that resembles the original object. The lens performs this task through the use of curved surfaces of materials with indices of refraction different than air, relying on the principle of refraction. Implicit in this description is the idea that light travels in only one general direction: from the object to the image. The astute student of optics, however, might detect a problem: The difference in index of refraction between two media (such as air and glass) that gives rise to refraction will also give rise to an unwanted phenomenon – reflection.

The design of a lens, including the ray tracing algorithms used to optimize the individual shapes and sizes of each glass or fused silica element in the lens, makes use of the assumption that all light traveling through the lens continues to travel from object to image *without any reflections*. How is this achieved in practice? One of the hidden technologies of lens manufacturing is the use of antireflection coatings on nearly every glass surface in a lens. These lens coatings are designed to maximize the transmittance of light at the interface between materials by creating an interference coating. These coatings, usually made of two layers, have specifically designed thicknesses and refractive indices which reduce reflections through interference among several reflected beams. These coatings are quite effective at reducing reflections at a specific wavelength and over a range of incident angles. This is critical, since a typical microlithographic lens today may have over 50 surfaces requiring coatings.

Although the lens coatings used today are quite good, they are not perfect. As a result, unwanted reflections, though small in magnitude, are inevitably causing light to bounce around within a lens. Eventually some of this light will make its way to the wafer. For the most part, these spurious reflections are reasonably random, resulting in a nearly uniform background light level exposing the wafer called *flare*. Flare is generally defined as the fraction of the total light energy reaching the wafer that comes from unwanted reflections and scatterings within the lens. As will be shown in the modeling section below, a more careful definition is required.

Lens coating non-ideality is not the only source of flare. Flare is caused by anything that causes the light to travel in a “non-ray trace” direction. In other words, reflections at an interface, scattering caused by particles or surface roughness, or scattering caused by glass inhomogeneity all result in stray light called flare (Figure 1). Defects such as these can be built into the lens during manufacturing, or can arise due to lens degradation (aging, contamination, etc.). The lithographic consequences of this stray light is probably obvious: degradation of image quality. Since flare is a nearly uniform background light level exposing the wafer, it provides exposure to nominally dark features (such as a line), reducing the image contrast and image log-slope.

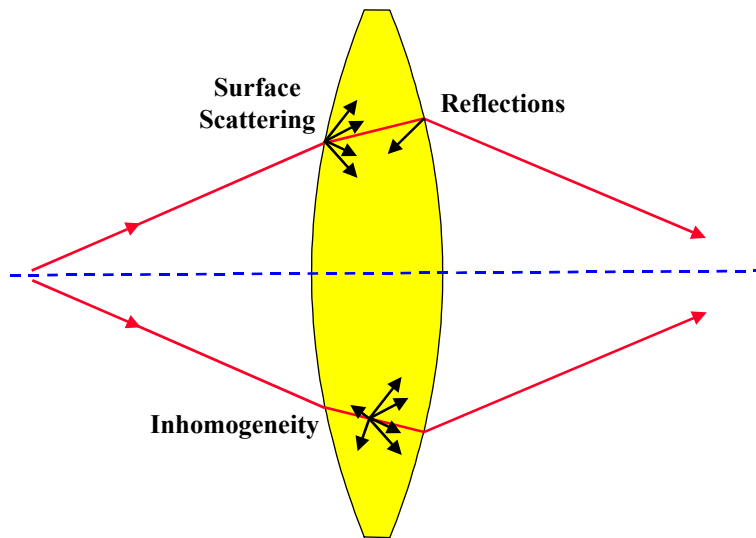


Figure 1. Flare is the result of unwanted scattering and reflections as light travels through an optical system. Shown here are the most common causes of random flare.

Besides the near-random scattered light described above, a second type of flare results in ghost images on the wafer. Reflections at the image or mask plane can send a secondary “ghost” image of the mask through the optical system in the wrong direction. A second reflection can send that image back to the wafer. Reflective surfaces such as the wafer, the backside of the chrome reticle, and lens surfaces all contribute to the formation of ghost images. Such images are very nearly eliminated if both the wafer and the backside of the mask absorber have very low reflectivity. In general, a good lithographic tool using a low reflectivity chrome mask and a wafer with a bottom

antireflective coating should have essentially no ghost imaging problems. Thus, random scatter flare is the focus of this work.

II. Measuring Flare – The Basics

A simple method for measuring flare was first proposed by Flagello and Pomerene [1]. Consider the imaging of an island feature whose dimension is extremely large compared to the resolution limits of the imaging tool (say, a $100\mu\text{m}$ square island in positive resist). In the absence of flare, the imaging of such a large feature will result in very nearly zero light energy at the center of the image of the island. The presence of flare, on the other hand, will provide light to this otherwise dark region on the wafer (Figure 2). A positive photoresist can be used as a very sensitive detector for low levels of flare.

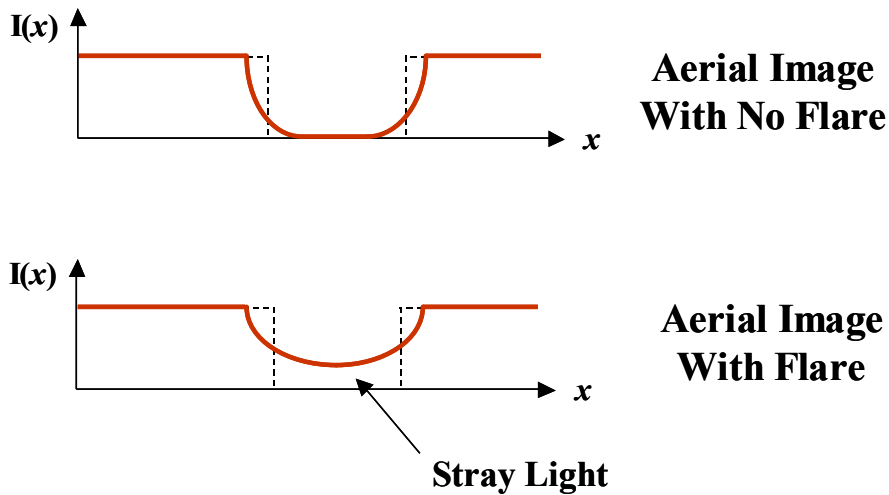


Figure 2. Cartoon plots of the aerial image intensity $I(x)$ for a large island mask pattern with and without flare.

The dose to clear (E_o) is defined as the minimum dose required to completely remove the photoresist during development for a large open frame exposure. A related concept is the island dose to clear ($E_o - \text{island}$), the minimum dose required to completely wash away a large island structure during a normal development process. In the absence of flare, a large island would take nearly an infinite dose to produce enough exposure at its center to make the resist soluble in developer. With flare present, however, this dose is reduced considerably. In fact, by measuring the normal dose to clear and the dose to clear for the large island, the amount of flare can be determined as

$$Flare = \frac{E_0}{E_{0-island}} \quad (1)$$

For example, if the dose to clear of a resist is 10 mJ/cm^2 , then an imaging tool with 5% flare would mean that a large island will clear with a dose of about 200 mJ/cm^2 . Note that one of the most important characteristics of this test is that by using the ratio of two “dose-to-clear” measurements, the influence of the photoresist (absorption, finite resist contrast, etc.) is completely normalized out. Any resist or resist process should produce the same measurement value for flare. Full contrast curves for both the clear field and the island can also be used to improve the accuracy of the measurement [2].

Although flare is a characteristic of an imaging tool, it is also a function of how that tool is used. For example, the amount of flare experienced by any given feature is a function of both the local environment around that feature (short range flare) and the total amount of energy going through the lens (long range flare) [2]. A darkfield reticle produces images with almost no flare, whereas a reticle which is almost 100% clear will result in the maximum possible flare. The data in Figure 3, taken from reference [2], shows the two distinct regions clearly. Here, a clearfield reticle was used with one flare target placed in the middle of the field. Framing blades were used to change the total clear area of the exposure field. As can be seen, flare quickly rises to a level of about 1%, then grows approximately linearly with clear area of the field. The y-intercept of the linear portion of the curve can be thought of as the short range contribution to the total flare.

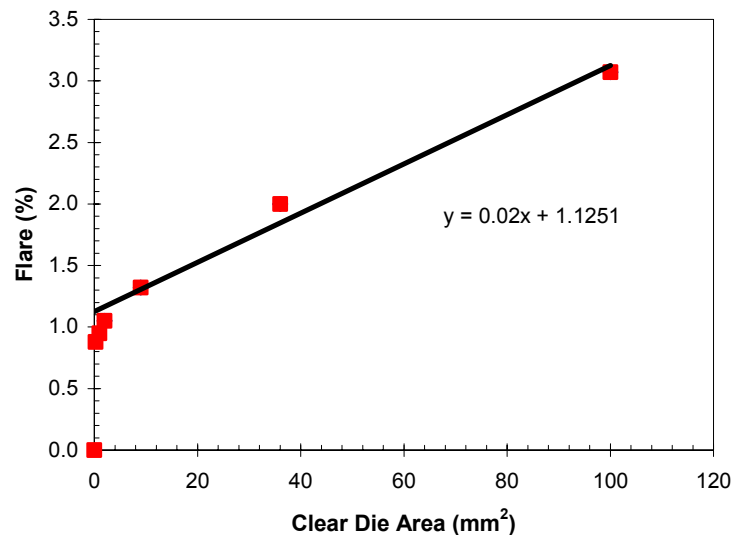


Figure 3. Using framing blades to change the field size (and thus total clear area of the reticle), flare was measured at the center of the field (from Ref [2]).

Flare is also a function of field position, with points in the center of the field often experiencing flare levels 50% higher than points near the edge of the field [3]. This phenomenon may be thought of as a side effect of the long range versus short range scattering discussed above. At the edge of the field, the local clear area is half of that at the center of the field.

II. Measuring Flare – The Details

Although measurement of the long range flare (also called DC flare) is straightforward and accurate, how can the more interesting phenomenon of short range flare be tested? In this section, we'll review a number of proposals for measuring this short range flare [2-7]. Besides the framing blade test described above, one obvious approach would be to vary the size of the flare pad and to look for any variation of the measured flare. Obviously, the center of a 100 μm square pad is locally dark for a region of 10,000 μm^2 , whereas a 50 μm square pad in an otherwise clear field would be only 25% dark in the same 10,000 μm^2 region. There is a limit, however, as to how small the flare pad can be made. The use of a "large" pad is required to ensure that only flare is measured and not diffraction or lateral development effects.

If an accuracy of 0.2% flare is required (that is, the ability to differentiate between 3% flare and 3.2% flare), then the contribution of light into the middle of the pad by diffraction would have to be less than 0.002. Assuming that the middle 500nm of the pad will be used to monitor the pad clearing dose, a simulation of the imaging (NA = 0.75, λ = 248nm) of an isolated pad using PROLITH v7.2 was carried out for conventional and annular illumination over a range of sigma values. Focus was allowed to vary over a range of $\pm 0.3\mu\text{m}$ and 30 mwaves of 3rd order coma and spherical aberrations were assumed to make the conditions somewhat worst-case, though still realistic. For the isolated pad with conventional illumination a pad size of 2 μm (k_1 = 6) was required to keep diffraction effects below the 0.002 intensity level, whereas for annular illumination a slightly larger pad, k_1 = 7, was needed. In all cases, the higher sigma values proved to be worst case. Results remained approximately the same when isolated line features were used instead of pads.

Lateral development can also cause the line or pad size to shrink, possibly making the feature disappear prematurely. In the central region of uniform flare exposure, the size of the pad must be such that the vertical development reaches the bottom of the center of the pad before the lateral development. Thus, the pad must be at least twice as wide as the resist thickness. Assuming a typical resist thickness of k_1 = 1.5 (using the normalized distance units for convenience), and since the aerial image edge region covers a length of about $0.5\lambda/\text{NA}$, this would mean the pad size would have to be greater than about k_1 = 4 or 5. Thus, it seems likely that diffraction effects will limit the smallest possible pad size before lateral development effects would become significant.

Luce performed a test varying the pad size [3], showing flare increasing from 6.8% for a 400 μm pad, to 7% for a 10 μm pad, to about 7.5% for a 5 μm pad, and finally to 8% for a 2 μm pad (corresponding to k_1 = 4.5 on the stepper used). Since the last data point should show some diffraction effects confounding the flare measurements, one could conclude that flare contributed

from an area farther away than $\pm 200\mu\text{m}$ amounted to 6.8%, while the shorter range flare coming from an area of $400 \times 400\mu\text{m}^2$ around the feature contributed another 1% approximately (Figure 4). Interestingly, this test indicates around 1% short range flare for this lens, similar to the value shown above in Figure 3 for a different lens at a different wavelength.

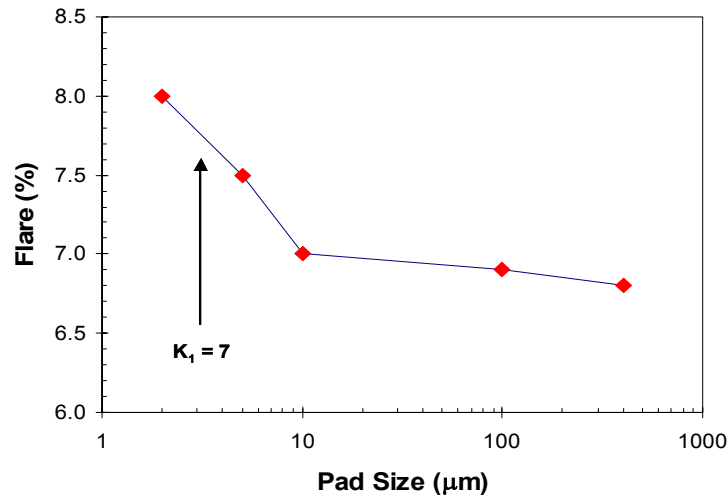


Figure 4. Varying measured flare with pad size shows the presence of some amount of short range flare (from Ref [3]).

Kirk has proposed an edge measurement technique as a way to extract the contribution of short range flare without the need to create a reticle with varying pad sizes [4]. In the Kirk receding edge method, one large pad is over exposed as for the standard flare test, but using doses lower than that needed to clear the pad. The change in edge position of one side of the pad is used as a measure of the flare at that specific location. The dose used, as a ratio to the open frame dose to clear, is then said to be the flare at this edge position. Unlike the standard flare measurement, however, the Kirk receding edge method assumes that the photoresist has infinite contrast (if a point in the resist receives a dose greater than E_0 , it is completely developed away). And although Kirk was careful to only make measurements past the diffraction edge region, lateral development effects will always be present since the development front at the edge of the pad will always be traveling laterally. Thus, interpretation of receding edge data is difficult at best, and would probably require simulation to properly extract flare values.

To test out the efficacy of the Kirk receding edge test, PROLITH v7.2 was used to simulate the experiment using only DC flare (i.e., a uniform level of flare independent of position or clear area). For a standard 248nm process assuming 500nm of UV6 resist on BARC, $NA = 0.75$, $\sigma = 0.5$, a large pad was simulated through exposure in the presence of 2% and 5% DC flare. For this case, no aberrations were assumed and best focus was used, giving a near ideal experimental condition. To simulate a less than ideal case, $0.3\mu\text{m}$ of defocus was added to the simulation. Analysis of the resulting edge position data provided the measured “flare” results in Figure 5. Note that past the

edge position of about 500nm, the pad completely washed away (i.e., the dose required to move the edge position to 500nm was enough to cause the DC flare to wash away the pad through vertical development). By interpreting the results of this test as a “localized flare” as a function of position, one would come to the entirely false conclusion that there is a large (5-10%) local flare component that operates over a range of less than 1 micron when in fact lateral development effects alone can explain the observed behavior.

Of course, this simulation exercise does not prove that the Kirk test cannot measure local flare. It simply shows that without careful attention to the details of the experiment it can be quite easy to misinterpret the results. Simulation is a powerful tool for testing the assumptions of an experimental analysis. For example, after extracting the parameters for a local flare model from experiment, simulation using that local flare model and parameters should be capable of reproducing the original experiment. At a minimum this provides a self-consistency check that should be routinely performed whenever lithographic parameters are extracted from the interpretation of experimental results. More work in this area for local flare measurements is certainly warranted.

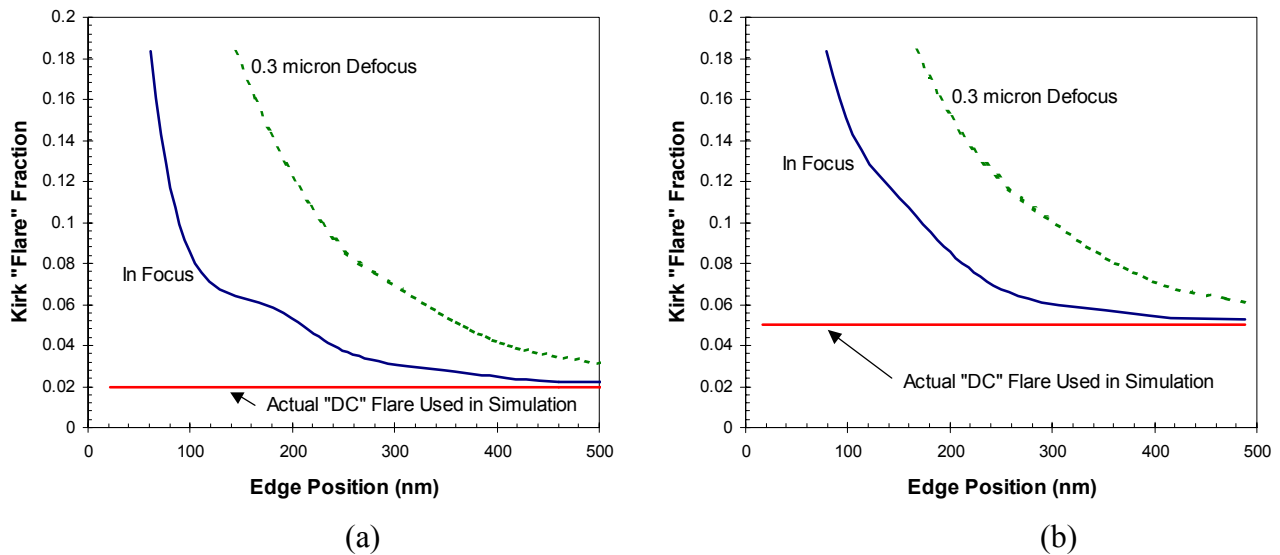


Figure 5. Simulation of the Kirk receding edge flare measurement method assuming (a) 2% DC flare and (b) 5% DC flare shows that the lateral development of a finite contrast resist creates a response that can be incorrectly interpreted as a “localized flare” response.

A double exposure method for flare measurement is also quite promising [5,6]. Here, a critical dimension feature is exposed and then a second exposure is performed where the critical feature is covered by a large pad. Without flare, the second exposure would have no impact on the developed critical feature width (assuming that the large blocking pad fulfills the requirements for less than 0.2% diffraction contribution). With flare, the second exposure provides a background

dose that will cause the critical feature to be overexposed and thus have a reduced critical dimension (CD). To calibrate this measurement, the impact of a blanket exposure on the CD of the small feature must first be measured. For this calibration, the first exposure is performed as normal, but the second exposure is performed with a clear reticle, creating a blanket exposure. The CD of the developed feature as a function of the second clear exposure dose becomes the calibration curve used for the flare measurement. However, since the same size restrictions exist for the blocking exposure as for the standard flare test, it is not clear that the double exposure method offers any clear advantage over the Flagello-Pomerene pad test.

Perhaps the ultimate local flare test has yet to be performed. Using a small flare pad ($k_1 = 8$, for example), the local area surrounding the pad would be filled with moderate sized lines and spaces ($k_1 = 1$) with a duty ratio adjusted to create a certain local transmittance. This region might extend out $\pm 100\mu\text{m}$ around the pad. Beyond this local fill area, the mask would be perfectly clear and these flare pads would be spaced on 1mm centers or larger. Thus, the mask would be 96+% clear overall. The very near region surrounding the pad ($k_1 = 4$ on all sides of the pad, for example) would be left clear to prevent any possible optical and physical proximity effects. A reticle full of these pads could be used to create a wide range of local transmittances which could be used to characterize the local flare independent of other effects.

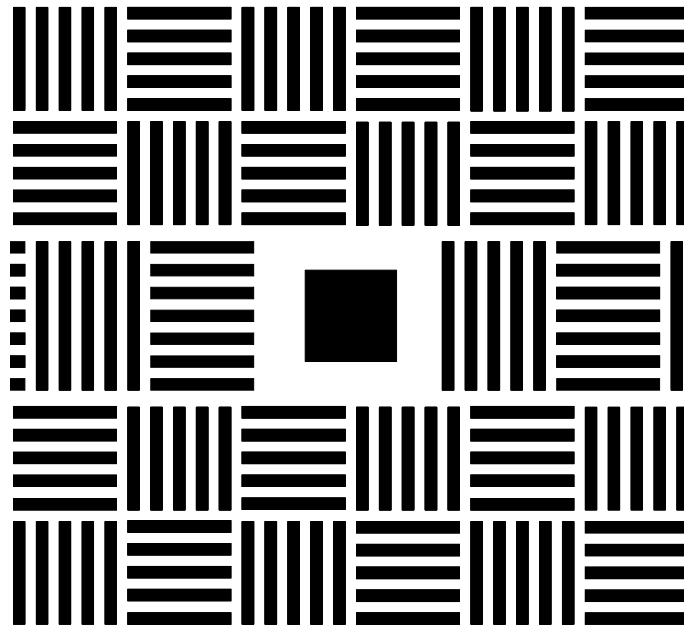


Figure 6. Proposed local flare test where the overall clear area of the reticle would be held constant at nearly 100% but the region within $\pm 100\text{mm}$ would have a variable transmittance controlled by the duty cycle of the line/space fill patterns.

III. Modeling Flare

Flare had been included in lithography simulators for many years, but only in a very simple form. DC (long range) flare is generally modeled by first calculating the image without flare, $I_o(x,y)$, and then adding flare as

$$I(x,y) = Flare + (1 - Flare)I_o(x,y) \quad (1)$$

where $Flare$ is a fractional value (0.03 for 3% flare, for example). In fact, even this simple DC model for flare has its problems, as will be shown below.

Consider a simple mechanism for the generation of DC flare [8]. Light passing through the lens is scattered by one or more of the mechanisms shown in Figure 1. In its simplest manifestation, flare would cause all light to scatter in equal proportion regardless of its spatial frequency and field position (angle and position with which it enters the lens). In this case, scattering will result in a uniform reduction in the energy that coherently interacts to form the image. If SF represents the *scatter fraction*, the fraction of the light energy (or intensity) that scatters and thus does not contribute to the coherent creation of the image, the resulting image would be $(I-SF)I_o(x,y)$. But where does this scattered light go? Another simple approximation for uniform DC flare would be that the scattered light is uniformly spread over the exposure field. Thus, a DC or background dose would be added to the aerial image. But while scattering reduces the intensity of the image *locally*, the background dose is added *globally*. Thus, the additional background dose would be equal to the scatter fraction multiplied by the total energy passing through the lens. A more accurate DC flare model would be

$$I(x,y) = SF \cdot EF + (1 - SF)I_o(x,y) \quad (2)$$

where EF is the energy fraction of the reticle, the ratio of the total energy reaching the wafer for this reticle compared to the energy that would reach the wafer for a perfectly clear reticle. As a reasonable approximation, the total energy passing through the lens is proportional to the clear area of the reticle [9]. Thus, a simple but still quite accurate DC flare model would be

$$I(x,y) = SF \cdot CA + (1 - SF)I_o(x,y) \quad (3)$$

where CA is the clear area fraction for the field. This new flare model corrects a significant deficiency in the original DC flare model: it obeys conservation of energy.

For a 100% clear area full field mask, equations (1) and (3) become the same. Thus, the standard measurement of flare for a clear field reticle would yield the scatter fraction required for the new DC flare model. Since the clear area fraction is generally known (at least approximately) for any mask, the new DC flare model still requires only one measurable parameter, the scatter fraction. For a dark field mask, as the clear area fraction goes to zero, the effect of flare is just a loss of dose for the image. It is doubtful that this dose loss is observable in normal lithographic practice since dark field exposure layers (such as contacts and vias) generally operate at different

numerical aperture/partial coherence combinations than clear field reticles, and dose calibration is not consistent across different stepper lens and illumination settings. The biggest difference between equations (1) and (3) when predicting the effects of flare would be for 50% clear area reticles. Here, flare in equation (3) would cause the full loss of peak intensity due to scattering, but would predict less background dose in the nominally dark regions than the old DC flare model.

A slight correction to equation (3) would occur if some fraction of the scattered light was assumed to be lost (absorbed into the lens housing, for example). However, for the low levels of flare expected for normal lithographic imaging tools (less than 10%), the effect of some of the scatter fraction being lost and not reaching the wafer would be a simple dose recalibration and would not be observable.

Equation (3) is an improved DC flare model, but still does not take into account the impact of short range flare effects, as observed in Figures 3 and 4. The scattering point spread function approach [4,6] will probably prove useful, though as discussed, many of the past attempts to measure this scattering point spread function have suffered from the experimental uncertainties associated with the receding edge technique.

IV. Conclusions

The measurement of flare by several common techniques have been reviewed. It seems that the original Flagello and Pomerene technique of the large flare pad is still the best, so long as the flare pad is kept larger than about $k_1 = 6$ or 7. By varying the global clear area (using framing blades, for example), the intercept of a linear flare versus clear area plot indicates the total amount of local flare versus long range (DC) flare. Either by changing the local clear area through fill patterns or by varying the pad size, the impact of short range flare can, in principle, be characterized.

A new, yet still quite simple, model for DC flare has been proposed that should exhibit improved accuracy. It has the added advantage of being capable of accounting for the impact of reticle clear area on the DC flare and has the comforting feature that it obeys the law of conservation of energy.

References

1. D. G. Flagello and A. T. S. Pomerene, "Practical Characterization of 0.5 μ m Optical Lithography," *Optical Microlithography VI, Proc.*, SPIE Vol. 772 (1987) pp. 6-20.
2. C. A. Mack and P. M. Kaufman, "Mask Bias in Submicron Optical Lithography," *Journal of Vacuum Science and Technology*, Vol. B6, No. 6 (Nov./Dec. 1988) pp. 2213-2220.
3. E. Luce, et al., "Flare Impact on the Intrafield CD Control for Sub-0.25 μ m Patterning," *Optical Microlithography XII, Proc.*, SPIE Vol. 772 (1999) pp. 368-381.

4. J. P. Kirk, "Scattered Light in Photolithographic Lenses," *Optical/Laser Microlithography VII, Proc.*, SPIE Vol. 2197 (1994) pp. 566-572.
5. A. Bourov, L. Litt, and L. Zavyalova, "Impact of Flare on CD Variation for 248nm and 193nm Lithography Systems," *Optical Microlithography XIV, Proc.*, SPIE Vol. 3679 (2001) pp. 1388-1393.
6. K. Lai, C. Wu, and C. Progler, "Scattered Light: The Increasing Problem for 193nm Exposure Tools and Beyond," *Optical Microlithography XIV, Proc.*, SPIE Vol. 772 (2001) pp. 1424-1435.
7. D. Nam, G. Yeo, E. Lee, H. Kim, M-H. Jung, H. Cho, W. Han, and J. Moon, "Is there a Royal Road to Standardization for Measuring and Monitoring Flare Effect," preprint of paper submitted to *Journal of Vacuum Science and Technology*.
8. C. A. Mack, "Measuring and Modeling Flare in Optical Lithography," *Arch Microlithography Symposium Interface 2002*, (Oct. 2002).
9. Because of diffraction, only a portion of the energy passing through the reticle actually passes through the lens. For example, for a mask pattern of equal lines and spaces where only the zero and ± 1 first orders pass through the lens, the energy going through the lens is about 90% of the energy that passes through the mask.