

Journal of
Micro/Nanolithography,
MEMS, and MOEMS

Nanolithography.SPIEDigitalLibrary.org

Reducing roughness in extreme ultraviolet lithography

Chris A. Mack

Reducing roughness in extreme ultraviolet lithography

Chris A. Mack*

Fractilia, LLC, Austin, Texas, United States

Abstract. Pattern roughness is a major problem in advanced lithography for semiconductor manufacturing, especially for the insertion of extreme ultraviolet (EUV) lithography as proposed in the coming years. Current approaches to roughness reduction have not yielded the desired results. Here, a global optimization approach is proposed, taking advantage of the different strengths and weaknesses of lithography and etch. Lithography should focus on low-frequency roughness by minimizing both the low-frequency power spectral density (PSD) and the correlation length. Etch should focus on high frequency roughness by growing the correlation length. By making unbiased measurements of the roughness, including the PSD, the parameters needed to guide these optimization efforts become available. The old approach, of individually seeking to reduce the 3σ roughness of pre- and postetch features, is unlikely to lead to the required progress in overall roughness reduction for EUV. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMM.17.4.041006](https://doi.org/10.1117/1.JMM.17.4.041006)]

Keywords: line-edge roughness; linewidth roughness; stochastic-induced roughness; line-edge roughness; linewidth roughness; power spectral density; power spectral density.

Paper 18052SS received May 3, 2018; accepted for publication Jul. 18, 2018; published online Aug. 7, 2018.

1 Introduction

Stochastic-induced roughness continues to be a major concern in the implementation of extreme ultraviolet (EUV) lithography for semiconductor high-volume manufacturing, potentially limiting product yield or lithography throughput or both. For this reason, considerable effort has been made in the last 10 years to characterize, understand, and reduce stochastic-induced roughness of postlithography and post-etch features. Despite these efforts, far too little progress has been made in reducing the effects of stochastics, such as linewidth roughness (LWR), line-edge roughness (LER), and local critical dimension uniformity (LCDU).¹

Reducing roughness requires a thorough understanding of roughness and its causes.^{2,3} And understanding roughness requires, among other things, trustworthy measurements of roughness. Further, roughness measurement must include frequency characterization in order to understand fully the nature of the roughness behavior at various length scales. This paper will begin by reviewing the frequency characterization of roughness using the power spectral density (PSD), then describe how to make unbiased measurements of the PSD (where noise coming from the SEM imaging is subtracted out). Finally, a simple model of roughness that makes use of the unbiased PSD will be presented. This model, and further insights about the role of etch processes in modifying the roughness coming from lithography, will lead to important conclusions about resist and etch process design for reduced roughness of the after-etch features.⁴

2 Frequency Dependence of Roughness

Rough features are most commonly characterized by the standard deviation of the edge position (for LER), linewidth (for LWR), or feature centerline for pattern placement roughness (PPR). But describing the standard deviation is not enough to fully describe the roughness. Figure 1 shows

four different rough edges, all with the same standard deviation. The obvious differences visible in the edges make it clear that the standard deviation is not enough to fully characterize the roughness. Instead, a frequency analysis of the roughness is required.

The standard deviation of a rough edge describes its variation relative to and perpendicular to an ideal straight line. In Fig. 1, the standard deviation describes the vertical variation of the edge. But the variation can be spread out differently along the length of the line (in the horizontal direction in Fig. 1). This line-length dependence can be described using a correlation function such as the autocorrelation function or the height–height correlation function. Alternatively, the frequency f can be defined as one over a length along the line (Fig. 2). The dependency of the roughness on frequency can be characterized using the PSD. The PSD is the variance of the edge per unit frequency (Fig. 2) and is calculated as the square of the coefficients of the Fourier transform of the edge deviation. The low-frequency region of the PSD curve describes edge deviations that occur over long length scales, whereas the high-frequency region describes edge deviations over short length scales. Commonly, PSDs are plotted on a log–log scale.

The PSD of lithographically defined features generally has a shape similar to that shown in Fig. 2. The low-frequency region of the PSD is flat (so-called “white noise” behavior), then above a certain frequency it falls off as a power of the frequency (a statistically fractal behavior). The difference in these two regions has to do with correlations along the length of the feature. Points along the edge that are far apart are uncorrelated with each other (statistically independent), and uncorrelated noise has a flat PSD. But at short length scales, the edge deviations become correlated, reflecting a correlating mechanism in the generation of the roughness, such as acid reaction-diffusion for a chemically amplified resist.⁵ The transition between uncorrelated and correlated behaviors occurs at a distance called the correlation length. Note that the exact definition of

*Address all correspondence to: Chris A. Mack, E-mail: chris@Lithoguru.com

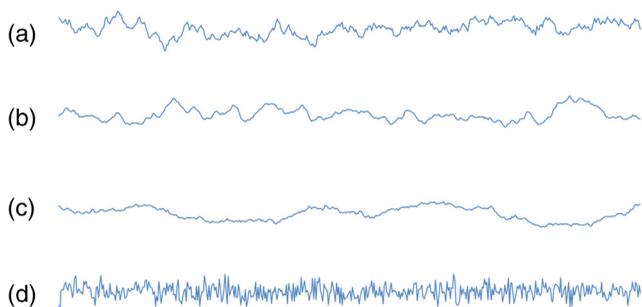


Fig. 1 These four randomly rough edges all have the same standard deviation of roughness, but differ in the frequency parameters of correlation length (ξ) and roughness exponent (H): (a) $\xi = 10$, $H = 0.5$, (b) $\xi = 10$, $H = 1.0$, (c) $\xi = 100$, $H = 0.5$, and (d) $\xi = 0.1$, $H = 0.5$.

the correlation length is arbitrary to within a multiplicative constant.⁵

Figure 3 shows that a typical PSD curve can be described with three parameters. PSD(0) is the zero frequency value of the PSD. While this value of the PSD can never be directly measured (zero frequency corresponds to an infinitely long line), PSD(0) can be thought of as the value of the PSD in the flat low-frequency region. The PSD begins to fall at a frequency of $1/(2\pi\xi)$, where ξ is the correlation length. In the fractal region, we have what is sometimes called “ $1/f$ ” noise and the PSD has a slope (on the log–log plot) corresponding to a power of $1/f$. The slope is defined as $2H + 1$, where H is called the roughness exponent (or Hurst exponent). For example, $H = 0.5$ for a purely reaction-diffusion process causing the correlation.^{5,6} Each of the parameters of the PSD curve has important physical meaning for a lithographically defined feature, and more about that meaning will be discussed in a subsequent section. The variance of the roughness is the area under the PSD curve and is derived from the other three PSD parameters.

A useful model for fitting the shape of a PSD curve was proposed by Palasantzas⁷ and has been used extensively to fit after-lithography and after-etch roughness results. A modified version of that model, however, has proven to be more useful in my experience

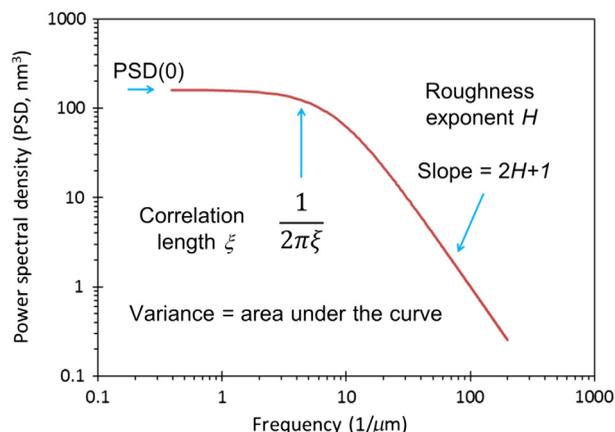


Fig. 3 A typical PSD can be described by three parameters: PSD(0), the low-frequency value of the PSD, the correlation length ξ , and the roughness exponent H . The variance of the roughness is the area under the PSD curve.

$$\text{PSD}(f) = \frac{\text{PSD}(0)}{1 + |2\pi f \xi|^{2H+1}} \tag{1}$$

The exact relationship between variance and the other three PSD parameters depends on the exact shape of the PSD curve in the midfrequency region (defined by the correlation length), but an approximate relationship based on Eq. (1) shows the general trend

$$\sigma^2 = \frac{\text{PSD}(0)}{(1.2H + 1.4)\xi} \tag{2}$$

The differences observed in the four rough edges of Fig. 1 can now be easily seen as differences in the PSD behavior of the features. Figure 4 shows two PSDs, corresponding to edge (a) and edge (c) from Fig. 1. While the two edges have the same variance (the same area under the PSD curve), they have different values of PSD(0) and correlation length (in this case the roughness exponent was kept constant). As we shall see, the different PSD curves will result in different roughness behavior for lithographic features of finite length.

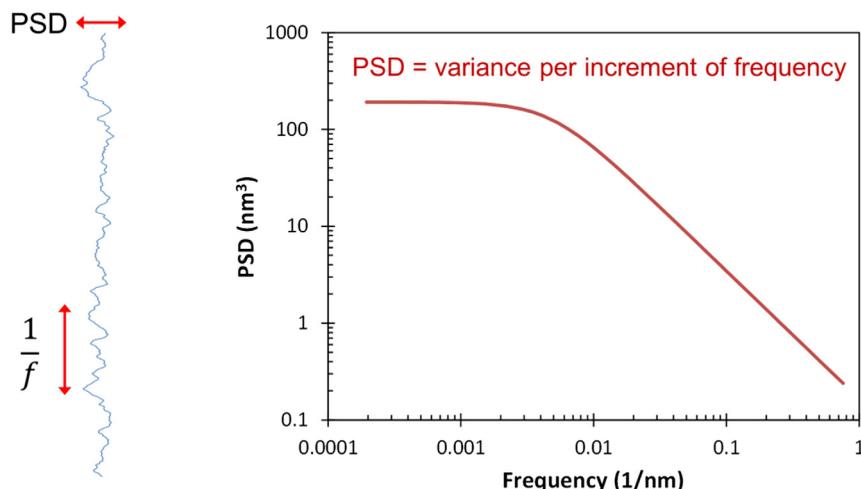


Fig. 2 An example of a rough edge and its corresponding PSD.

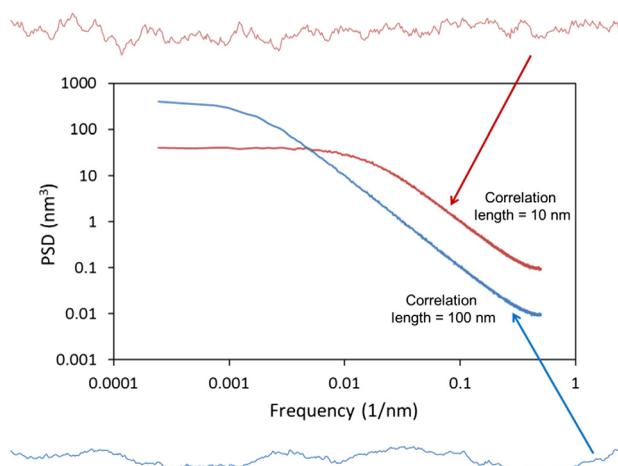


Fig. 4 Two edges from Fig. 1, edge (a) and edge (c), are shown to have different PSD behavior even though the standard deviations of the roughness are the same.

3 Device Impact of the Frequency Behavior of Roughness

The roughness of lines and spaces is characterized by measuring very long lines and spaces, long enough so that the flat region of the PSD becomes apparent. For a sufficiently long feature, the measured LWR can be thought of as the LWR of an infinitely long feature, $\sigma_{LWR}(\infty)$. But semiconductor devices are made from features that have a variety of lengths L . For these shorter features, stochastics will cause within-feature roughness, $\sigma_{LWR}(L)$, and feature-to-feature variation described by the standard deviation of the mean linewidths of the features, $\sigma_{CDU}(L)$. This feature-to-feature variation is called the local critical dimension uniformity, LCDU, since it represents CD variation that is not caused by the well-known “global” sources of error (scanner aberrations, mask illumination nonuniformity, hotplate temperature variation, etc.).⁸

For a line of length L , the within-feature variation and the feature-to-feature variation can be related to the LWR of an infinitely long line (of the same nominal CD and pitch) by the conservation of roughness principle⁹

$$\sigma_{CDU}^2(L) + \sigma_{LWR}^2(L) = \sigma_{LWR}^2(\infty). \quad (3)$$

The conservation of roughness principle says that the variance of a very long line is partitioned for a shorter line into within-feature variation and feature-to-feature variation. How this partition occurs is determined by the correlation length, or more correctly by L/ξ . Using a basic model for the shape of the PSD, we find that¹⁰

$$\sigma_{CDU}^2(L) = \frac{PSD(0)}{L} \left[1 - \frac{\xi}{L} (1 - e^{-L/\xi}) \right]. \quad (4)$$

Thus, Eqs. (1)–(3) show that a measurement of the PSD for a long line, and its description by the parameters $PSD(0)$, ξ , and H , enables one to predict the stochastic influence on a line of any length L . It is interesting to note that the LCDU does not depend on the roughness exponent, making H less important than $PSD(0)$ and ξ . For this reason, it is useful to describe the frequency dependence of roughness using an alternate triplet of parameters: $\sigma_{LWR}(\infty)$, $PSD(0)$, and ξ .

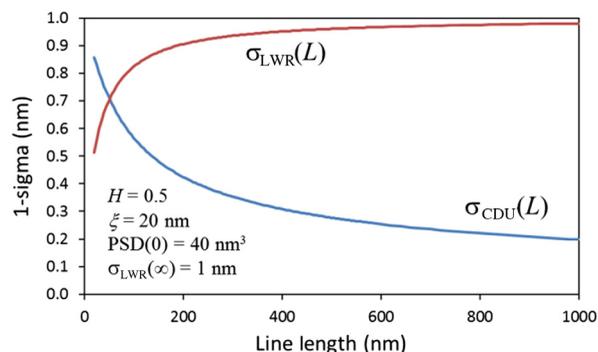


Fig. 5 The conservation of roughness principle showing how the within-feature roughness $\sigma_{LWR}(L)$ and the local CDU $\sigma_{CDU}(L)$ vary as a function of the line length L for a pattern of long lines and/or spaces.

Note that these same relationships apply to LER and PPR as well.

Examining Eq. (4), the correlation length is the length scale that determines whether a line of length L acts “long” or “short.” For a long line, $L \gg \xi$ and the local CDU behaves as

$$\sigma_{CDU}(L) \approx \sqrt{\frac{PSD(0)}{L}} \quad \text{when } L \gg \xi. \quad (5)$$

This long-line result provides a useful interpretation for $PSD(0)$: it is the square of the LCDU times the length of the line. Reducing $PSD(0)$ by a factor of 4 reduces the LCDU by a factor of 2, and the other PSD parameters have no impact (so long as $L \gg \xi$). Typically, resist has yielded correlation lengths on the order of one-third to one-half of the minimum half-pitch of their lithographic generation. Etch processes often increase the correlation length by 50% to 100%. Thus, when features are longer than about 5 to 10 times the minimum half-pitch of the technology node we are generally in this long line length regime. For shorter line lengths, the correlation length begins to matter as well.

Equations (3)–(5) show a trade-off of within-feature variation and feature-to-feature variation as a function of line length. Figure 5 shows an example. For very long lines, LCDU is small and within-feature roughness approaches its maximum value. For very short lines the LCDU dominates. However, due to the quadratic nature of the conservation of roughness, $\sigma_{LWR}(L)$ rises very quickly as L increases, but LCDU falls very slowly as L increases. Thus, there is a wide range of line lengths where both feature roughness and LCDU are significant.

4 Unbiased Measurement of PSD

By far the most common way to measure feature roughness is the top-down critical dimension scanning electron microscope (CD-SEM). CD-SEMs have been optimized for measuring mean critical dimension with high precision but have proven very useful for measuring LER, LWR, PPR, and their PSDs as well. However, some errors in the SEM images can have large impacts on the measured PSD while having almost no impact on the measurement of mean CD.¹¹ For this reason, the metrology approach needed for PSD

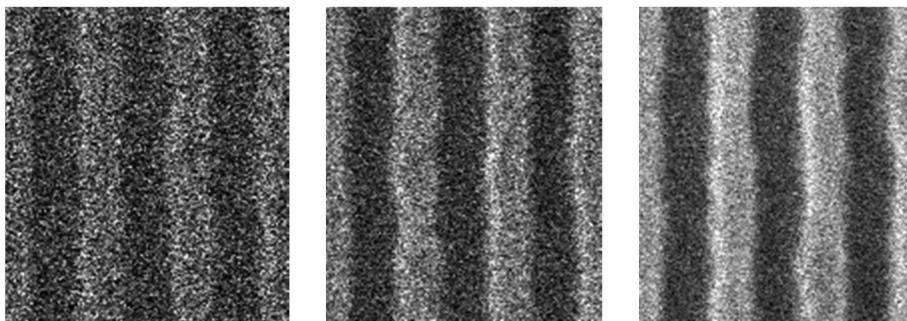


Fig. 6 Portions of SEM images of nominally identical resist features with 2, 8, and 32 frames of integration (respectively, from left to right). Doubling the frames of integration doubles the electron dose per pixel. Since the dose is increased by a factor of 4 in each case, the noise goes down by a factor of 2. (Images provided in collaboration with imec.)

measurement may be quite different than the approach commonly used for mean CD measurement.¹²

The biggest impediment to accurate roughness measurement is noise in the CD-SEM image. SEM images suffer from shot noise, where the number of electrons detected for a given pixel varies randomly. For the expected Poisson distribution, the variance in the number of electrons detected for a given pixel of the image is equal to the expected number of electrons detected for that pixel. Since the number of detected electrons is proportional to the number of electrons that impinge on that pixel, noise can be reduced by increasing the electron dose that the sample is subjected to. For some types of samples, electron dose can be increased with few consequences. But for other types of samples (especially photoresist), high electron dose leads to sample damage (resist line slimming, for example). Thus, to prevent sample damage electron dose is kept as low as possible, where the lowest dose possible is limited by the noise in the resulting image. Figure 6 shows portions of three SEM images of nominally the same lithographic features taken at different electron doses.

Making the very reasonable assumption that the amount of edge detection noise in a SEM is independent of the amount of actual roughness of the feature, SEM image noise adds to the roughness of the patterns on the wafer to produce a measured roughness that is biased higher¹³

$$\sigma_{\text{biased}}^2 = \sigma_{\text{unbiased}}^2 + \sigma_{\text{noise}}^2, \quad (6)$$

where σ_{biased} is the roughness measured directly from the SEM image, σ_{unbiased} is the unbiased roughness (that is, the true roughness of the wafer features), and σ_{noise} is the random error in detected edge position (or linewidth) due to noise in the SEM imaging. Since an unbiased estimate of the feature roughness is obviously what is desired, the measured roughness must be corrected by subtracting an estimate of the noise term.

While several approaches for estimating the SEM noise and subtracting it out have been proposed,^{13–17} these approaches have not proven successful for today's small feature sizes and high levels of SEM image noise. The problem is the lack of edge detection robustness in the presence of high image noise: when noise levels are high, edge detection algorithms often fail to find the edge. The solution to this problem is typically to filter the image, smoothing out the high frequency noise. For example, if a Gaussian 7×3 filter

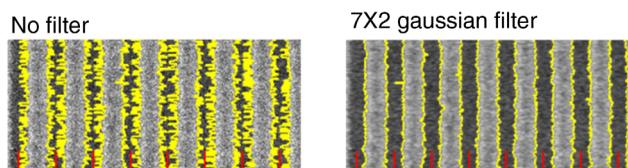


Fig. 7 Detecting edges in a noisy SEM image with and without the use of an image filter. From Ref. 18.

is applied to the image, then for each rectangular region of the image 7 pixels wide and 3 pixels tall, the grayscale values for each pixel are multiplied by a Gaussian weight and then averaged together. The result is assigned to the center pixel of the rectangle. This smoothing makes edge detection significantly more robust when image noise is high. Figure 7 shows an example of using a simple threshold edge detection algorithm with and without image filtering.¹⁸ Without image filtering, the edge detection algorithm is mostly detecting the noise in the image and does not reliably find the edge.

The use of image filtering can have a large effect on the resulting PSD. Figure 8 shows the impact of two different image filters on a collection of 30 images.¹⁸ All images were measured using an inverse linescan model for edge detection (as described later). Obviously the high-frequency region is greatly affected by filtering. But even the low-frequency region of the PSD shows a noticeable change when using a smoothing filter. Filtering in the y-direction

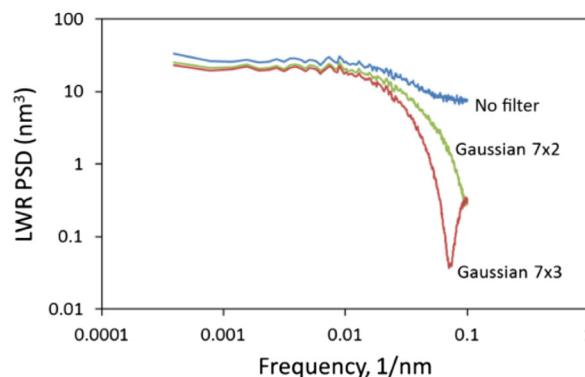


Fig. 8 Power spectral densities from many rough features with images preprocessed using a 7×2 or 7×3 Gaussian filter, or not filtered at all. From Ref. 18.

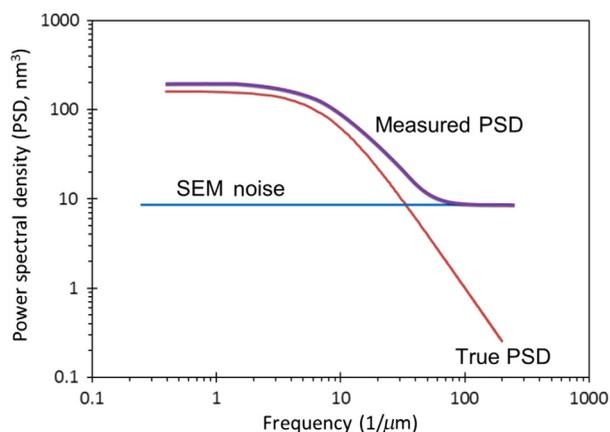


Fig. 9 The principle of noise subtraction: using the PSD, measure the flat noise floor in the high-frequency portion of the measured PSD, then subtract the white noise to get the true PSD.

throws away high-frequency information, whereas filtering in the x -direction lowers the linescan slope and can change the low-frequency behavior. As will be described next, the use of image filtering makes measurement and subtraction of image noise impossible.

If edge detection without image filtering can be accomplished, noise measurement and subtraction can be achieved by contrasting the PSD behavior of the noise with the PSD behavior of the actual wafer features. We expect resist features (as well as after-etch features) to have a PSD behavior as shown in Fig. 3. Correlations reduce high-frequency roughness so that the roughness becomes very small over very small length scales. SEM image noise, on the other hand, can be reasonably assumed to be white noise, so that the noise PSD is flat. Thus, at a high enough frequency, the measured PSD will be dominated by image noise and not actual feature roughness (the so-called “noise floor”).¹⁹ Given the grid size along the length of the line (Δy), SEM noise affects the PSD according to²⁰

$$\text{PSD}_{\text{biased}}(f) = \text{PSD}_{\text{unbiased}}(f) + \sigma_{\text{noise}}^2 \Delta y. \quad (7)$$

Thus, measurement of the high-frequency PSD (in the absence of any image filtering) provides a measurement of the SEM image noise. Figure 9 shows this approach. Clearly, this approach to noise subtraction cannot be used on PSDs coming from images that have been filtered since the filtering removes the high-frequency noise floor (see Fig. 8).

The key to using the above approach of noise subtraction for obtaining an unbiased PSD [and thus unbiased estimates of the parameters $\sigma_{\text{LWR}}(\infty)$, $\text{PSD}(0)$, and ξ] is to robustly detect edges without the use of image filtering. This can be accomplished using an inverse linescan model.¹⁸ A linescan model (such as the analytical linescan model^{21–23}) predicts the SEM image linescan given a set of beam conditions and the feature geometry on the wafer. Ideally, such a model would be physically based, easily calibrated, and not computationally intensive. An inverse linescan model runs this linescan model in reverse: given a measured linescan, what wafer feature edge positions produce a linescan that best fits the data? Such an inverse linescan model can use the physics of SEM image formation to constrain the possible

linescan shapes and reject the noise in the measured linescan to extract its signal. An inverse linescan model was used to generate the no-filter PSD data shown in Fig. 8.

Other SEM errors can influence the measurement of roughness PSD as well. For example, SEM field distortion can artificially increase the low-frequency PSD for LER and PPR, although it has little impact on LWR.¹¹ Background intensity variation in the SEM can also cause an increase in the measured low-frequency PSD, including LWR as well as LER and PPR. If these variations can be measured, they can potentially be subtracted out, producing the best possible unbiased estimate of the PSD and its parameters. As we will see in the following section, unbiased estimates of the PSD parameters can be used in models for stochastic-induced roughness, which in turn can be used to search for ways to reduce roughness.

5 Model for Stochastic-Induced Roughness in Lithography

A basic model for roughness has been proposed many times before: an error in the final resist edge position is equal to an error in the development rate R at the edge of the resist (position x) divided by the gradient in development rate^{19,24}

$$\Delta x = \frac{\Delta R}{dR/dx}. \quad (8)$$

For a random variation in development rate characterized by a mean and standard deviation, the resulting edge position will have a variation described by the 1-sigma LER

$$\sigma_{\text{LER}} = \frac{\sigma_R}{dR/dx}. \quad (9)$$

In this simple model, variation in the development path is ignored, which might be reasonable for small variations in development rate.^{25–28}

Development rate is determined by the level of remaining protecting groups (m) for a chemically amplified resist. This, in turn, is determined by the acid concentration (h) during a process of reaction-diffusion. Acid concentration is determined by the intensity of absorbed light (I_{abs}). In other words, an aerial image leads to an absorbed light image that leads to an acid latent image that leads to a protecting group latent image that leads to a development rate latent image. In a standard chemically amplified resist process, the only source of information about the correct position of the resist feature edge comes from the aerial image. Thus, at each step in this sequence, errors can increase the uncertainty (noise) and decrease the gradient (signal), making their ratio higher.^{29,30} This can be expressed as a propagation of noise/signal ratios

$$\sigma_{\text{LER}} = \frac{\sigma_R}{dR/dx} \geq \frac{\sigma_m}{dm/dx} \geq \frac{\sigma_h}{dh/dx} \geq \frac{\sigma_{I_{\text{abs}}}}{dI_{\text{abs}}/dx}. \quad (10)$$

The driver for LER is the last term in Eq. (10), which is also the minimum possible LER. Since the intensity of absorbed photons is proportional to the number of absorbed photons (N_{abs}), the minimum LER can also be expressed in terms of the number of photons absorbed at the line edge. Since the number of absorbed photons will follow a Poisson distribution

$$\sigma_{N_{\text{abs}}} = \sqrt{\langle N_{\text{abs}} \rangle}. \quad (11)$$

The aerial image log-slope (ILS) will equal the absorbed ILS for a nonbleaching resist so that

$$\text{ILS} = \frac{d \ln I}{dx} = \frac{1}{\langle N_{\text{abs}} \rangle} \frac{d \langle N_{\text{abs}} \rangle}{dx}. \quad (12)$$

This then gives an alternate expression for the smallest possible LER

$$\min \sigma_{\text{LER}} = \frac{\sigma_{I_{\text{abs}}}}{dI_{\text{abs}}/dx} = \frac{1}{\text{ILS} \sqrt{\langle N_{\text{abs}} \rangle}}. \quad (13)$$

The mean number of photons absorbed in some small volume of resist V is determined by the mean incident dose E (#photons/nm²) and the absorption coefficient α

$$\langle N_{\text{abs}} \rangle = \alpha V E. \quad (14)$$

As a numerical example, consider a volume that is a cube 10 nm on a side, a dose at the line edge of 6 photons/nm² (corresponding to 8.8 mJ/cm² of EUV light), an absorption coefficient of 0.007 nm⁻¹, and a normalized image log-slope (NILS) of 2 for a CD of 16 nm (ILS = NILS/CD). The minimum σ_{LER} will be 1.1 nm.

For the above expressions, everything is well known for a given lithographic case except the volume V . What is the correct ambit volume to average over? A smaller volume will produce a larger LER, so there must be some physical reason for the volume chosen. The smallest volume that might make sense is the size of one resist polymer molecule. After all, one molecule either dissolves or does not, and it is the sum of all the events that lead to dissolution that influence that dissolution. In general, however, the distance over which an absorbed photon might influence the dissolution of a resist molecule is larger than the size of the resist molecule. For a chemically amplified resist, an absorbed photon can lead to a generated acid which then diffuses some distance before causing a deprotection reaction, thus changing the solubility of the resist. The acid diffusion length, generally larger than the size of a resist polymer molecule, thus determines the volume of influence of an absorbed photon.

Put another way, all mechanisms that spread the influence of an absorbed photon through the resist determine the influence range and the ambit volume needed in Eq. (14). This spread is generally called the resist blur and includes not only acid diffusion but also secondary electron blur for an EUV resist. The ambit volume will then be proportional to the cube of the total resist blur.³¹ In addition, this influence range is also characterized by the resulting correlation length of the roughness, so the correlation length is a measure of the total resist blur. This means that

$$V \propto \xi^3. \quad (15)$$

Combining Eqs. (13)–(15) gives essentially Gallatin's classic LER model.¹⁹ The key insight here is the recognition that the correlation length of resist features is a measure of resist blur.

But blurring has another impact on lithography; it reduces the effective ILS and the gradient in the various latent

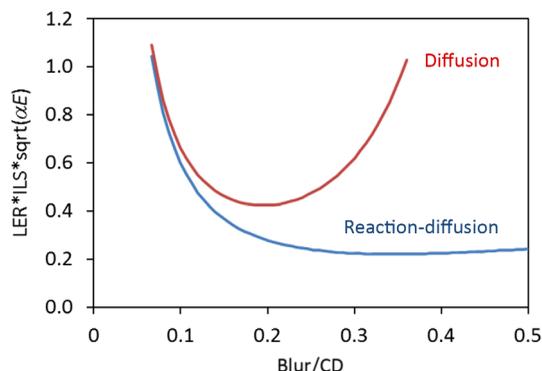


Fig. 10 Using the principles that the ambit volume and the effective ILS are both affected by the total resist blur (which is proportional to the correlation length of the roughness), there will be an optimum blur as a fraction of the nominal CD to produce minimum roughness.

images. Consider both a simple diffusion process (probably appropriate for secondary electron blur) and a reaction-diffusion process (appropriate for acid diffusion during postexposure bake). The reduction in the effective ILS has been previously derived for both cases²⁴

$$\text{Diffusion: } \frac{\partial \ln I_{\text{eff}}}{\partial x} \approx \frac{\partial \ln I}{\partial x} [e^{-2(\pi\xi/\text{CD})^2}],$$

$$\text{Reaction-diffusion: } \frac{\partial \ln I_{\text{eff}}}{\partial x} \approx \frac{\partial \ln I}{\partial x} \left[\frac{1 - e^{-2(\pi\xi/\text{CD})^2}}{2(\pi\xi/\text{CD})^2} \right], \quad (16)$$

where here the correlation length is assumed to be exactly equal to the diffusion length, though in fact there is likely some proportionality factor of order one, and CD is the half pitch for a pattern of small lines and spaces.

Replacing the ILS in Eq. (13) with the effective ILS, there will be an optimum correlation length balancing the competing factors of increasing the ambit volume and decreasing the effective ILS with larger ξ .³² Figure 10 shows that the optimum blur (correlation length) is about 20% of the half-pitch CD for the case of pure diffusion, and 35% of the half-pitch CD for the case of reaction-diffusion. As mentioned above, however, there may be a proportionality factor involved in the relationship between correlation length and diffusion length different from the proportionality factor involved in its use in the ambit value, so that we can only conclude that the optimum correlation length is some fraction of the minimum CD, probably in the 1/6 to 1/2 range.

If the total resist blur (correlation length) is optimized to produce the minimum roughness, that minimum roughness will scale as

$$\min \sigma_{\text{LER}} \propto \frac{1}{\text{NILS} \sqrt{\alpha E \text{CD}}}, \quad (17)$$

where the NILS is the aerial ILS multiplied by the nominal CD. This final result provides important scaling information about roughness. First, as many others have noted, roughness is inversely proportional to NILS. Since another important lithographic metric, exposure latitude, is also proportional to NILS, the long history of efforts in semiconductor lithography to improve NILS and exposure latitude have the added

benefit of reducing roughness. Unfortunately, the equally long history of living with lower NILS by reducing the sources of global variations (scanner aberrations, mask illumination nonuniformity, hotplate temperature variation, etc.) means that we are also living with higher roughness (since the sources of stochastic variation are not being reduced). Second, we can reduce the impact of photon shot noise by increasing the product of resist absorption coefficient and exposure dose.

Finally, optimizing the resist blur for minimum roughness at each new generation of critical dimension will result, other things being equal, in growing absolute roughness as feature size decreases. The relative roughness (roughness as a percentage of the nominal CD) will grow even faster. Since NILS is unlikely to increase as feature size decreases from one lithography generation to the next (the opposite is usually the case), this unpleasant aspect of roughness scaling means that exposure dose and/or absorption must grow inversely to CD to keep the absolute roughness constant. To keep the relative roughness constant from one lithography generation to the next, αE must be kept proportional to $1/CD^3$. If CD shrinks by 0.7, exposure dose must increase by a factor of 3 (all other things being equal) to keep the relative resist roughness constant.

6 Importance of Etch

The scaling result derived in the previous section only applies to the roughness of resist features. In semiconductor manufacturing, what is often most important is the roughness of the after-etch features. It is well known that etch reduces roughness, mostly through an increase in correlation length.³³ If this important feature of etch is combined with the scaling relationship for resist roughness above, an interesting opportunity arises. To keep roughness low, we must scale the postlithography correlation length in proportion to the CD. Further, current correlation lengths may in fact be larger than optimum so that even more reduction in correlation length could be helpful. But as Eq. (2) shows, a smaller correlation length leads to higher roughness for a given PSD(0). The difficulty comes from the coupling of correlation length and PSD(0) as is common in most resists and as described in the previous section. Higher correlation lengths mean larger resist blur, with a negative impact on latent image gradient and a corresponding increase in sensitivity to stochastic noise. Thus, PSD(0) and correlation length are generally not independent of each other.³⁴

Etch provides an important optimization opportunity since the growth in correlation length during etch comes with no equivalent trade-off in “blur.” For an etch process, PSD(0) and correlation length are not coupled. This leads to a new and important approach to minimizing the after-etch roughness. In lithography, we should optimize the resist and its process for both minimum PSD(0) and minimum ξ . This can be done without regard to minimizing the LER (σ_{LER} or σ_{LWR}) per se. In fact, a lithography process with minimum PSD(0) and minimum ξ will be unlikely to result in minimum postlithography roughness standard deviation. Then, we use the etch process to grow the correlation length, improving the high-frequency roughness that was ignored postlithography [while being sure not to worsen PSD(0), or lowering it if possible]. The final after-etch features will have minimum PSD(0), maximum correlation length,

and minimum σ_{LER} or σ_{LWR} . In other words, the lithography process should be made responsible for low-frequency roughness while the etch process is responsible for high-frequency roughness. This combination produces minimum roughness.

The proposed roughness optimization scheme involves a very different mindset than is often exhibited today. It is common today to “blame” the resist for roughness that is too high, then give credit to the etch process for “fixing” the roughness. It is also common today to attempt lithography optimization considering only the 3σ roughness as the metric to be reduced, ignoring the individual roles of PSD(0) and ξ . Further, lithography and etch processes are today typically optimized individually, without regard to how one influences the other. All of these ideas are flawed. Instead, lithography and etch should be optimized together, playing to the constraints and strengths of each process to individually optimize 3σ , PSD(0), and ξ . Several recent efforts have begun to prove out the worth of this idea.^{34,35} It is worth noting that the discussion so far has focused on resists and their influence on roughness. For EUV lithography, underlayers interact with the resist (e.g., by contributing secondary electrons during exposure) in a complicated way.³⁵

7 Conclusions

Reducing roughness in EUV lithography is extremely important and also extremely difficult without fairly large increases in exposure dose. In this paper, I have outlined a new strategy for optimizing the after-etch roughness of features by employing a synergy between etch and lithography. Lithography should focus on low-frequency LER by minimizing both PSD(0) and correlation length (a consequence of the coupled nature of these two parameters for lithographic features), or at least by minimizing PSD(0) without regard to correlation length. This optimization may not result in the lowest possible 3σ roughness for lithographic features. The etch process is then employed to minimize PSD(0) and maximize correlation length (a consequence of the uncoupled nature of these two parameters for after-etch features). Thus, etch is focused on improving the high-frequency roughness that lithography should ignore. The result should be a global optimum not obtainable by separately optimizing lithographic and etched features for 3σ roughness. This optimization scheme makes use of the insight that the correlation length of resist features is a measure of total resist blur.

Of course, in any regime where photon shot noise is an important component of overall roughness, increasing the dose is very effective at reducing roughness, though costly in a regime of low source intensity. Another effective way to increase the number of photons used to print a space without increasing the dose is to use phase-shifting masks. For example, a switch to the equivalent of a “chromeless” phase shifting mask for a pattern of equal lines and spaces is the same as doubling the exposure dose since the mask uses more of the photons to form the image rather than absorbing them. For contact holes, something like a factor of four increase in mask efficiency is possible.³⁶ While an absorberless EUV phase shifting mask will be difficult to make and control, it will likely be less difficult than another doubling or quadrupling of the intensity of the EUV light source.

The proposed litho + etch roughness reduction approach requires accurate measurement of unbiased values of $\sigma_{LWR}(\infty)$, PSD(0), and ξ . Relying solely on $\sigma_{LWR}(\infty)$, and especially its biased measurement, will be unlikely to produce the information needed to guide resist, resist process, etch tool, and etch process improvement.

References

1. Y.-J. Fan et al., "Benchmarking study of EUV resists for NXE:3300B," *Proc. SPIE* **9776**, 97760W (2016).
2. S. G. Hansen, "Photoresist and stochastic modeling," *J. Micro/Nanolithogr. MEMS MOEMS* **17**(1), 013506 (2018).
3. P. De Bisschop, "Stochastic effects in EUV lithography: random, local CD variability, and printing failures," *J. Micro/Nanolithogr. MEMS MOEMS* **16**(4), 041013 (2017).
4. C. A. Mack, "Reducing roughness in extreme ultraviolet lithography," *Proc. SPIE* **10450**, 10450OP (2017).
5. C. A. Mack, "Reaction-diffusion power spectral density," *J. Micro/Nanolithogr. MEMS MOEMS* **11**(4), 043007 (2012).
6. C. A. Mack, "Stochastic modeling in lithography: autocorrelation behavior of catalytic reaction-diffusion systems," *J. Micro/Nanolithogr. MEMS MOEMS* **8**(2), 029701 (2009).
7. G. Palasantzas, "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," *Phys. Rev. B* **48**(19), 14472-14478 (1993).
8. J. Finders et al., "Contrast optimization for 0.33NA EUV lithography," *Proc. SPIE* **9776**, 97761P (2016).
9. G. F. Lorusso et al., "Spectral analysis of line width roughness and its application to immersion lithography," *J. Micro/Nanolithogr. MEMS MOEMS* **5**(3), 033003 (2006).
10. C. A. Mack, "Analytical expression for impact of linewidth roughness on critical dimension uniformity," *J. Micro/Nanolithogr. MEMS MOEMS* **13**(2), 020501 (2014).
11. B. Lane et al., "Global minimization line-edge roughness analysis of top down SEM images," *Proc. SPIE* **10145**, 101450Y (2017).
12. G. Lorusso et al., "The need for line-edge roughness metrology standardization: the imec protocol," *Proc. SPIE* **10585**, 105850D (2018).
13. J. S. Villarrubia and B. D. Bunday, "Unbiased estimation of linewidth roughness," *Proc. SPIE* **5752**, 480 (2005).
14. A. Yamaguchi et al., "Bias-free measurement of LER/LWR with low damage of CD-SEM," *Proc. SPIE* **6152**, 61522D (2006).
15. R. Katz et al., "Bias reduction in roughness measurement through SEM noise removal," *Proc. SPIE* **6152**, 61524L (2006).
16. A. Yamaguchi et al., "Single-shot method for bias-free LER/LWR evaluation with little damage," *Microelectron. Eng.* **84**(5-8), 1779-1782 (2007).
17. S.-B. Wang et al., "Practical and bias-free LWR measurement by CDSEM," *Proc. SPIE* **6922**, 692222 (2008).
18. C. A. Mack and B. D. Bunday, "Using the analytical linescan model for SEM metrology," *Proc. SPIE* **10145**, 101451R (2017).
19. G. Gallatin, "Resist blur and line edge roughness," *Proc. SPIE* **5754**, 38-52 (2005).
20. C. A. Mack, "Systematic errors in the measurement of power spectral density," *J. Micro/Nanolithogr. MEMS MOEMS* **12**(3), 033016 (2013).
21. B. D. Bunday and C. A. Mack, "Influence of metrology error in measurement of line edge roughness power spectral density," *Proc. SPIE* **9050**, 90500G (2014).
22. C. A. Mack and B. D. Bunday, "Analytical linescan model for SEM metrology," *Proc. SPIE* **9424**, 94240F (2015).
23. C. A. Mack and B. D. Bunday, "Improvements to the analytical linescan model for SEM Metrology," *Proc. SPIE* **9778**, 97780A (2016).
24. C. A. Mack, *Fundamental Principles of Optical Lithography*, Chapter 9, John Wiley and Sons, London (2007).
25. A. R. Neureuther and C. G. Willson, "Reduction in x-ray lithography shot noise exposure limit by dissolution phenomena," *J. Vac. Sci. Technol. B* **6**(1), 167-173 (1988).
26. C. A. Mack, "Stochastic modeling in lithography: use of dynamical scaling in photoresist development," *J. Micro/Nanolithogr. MEMS MOEMS* **8**(3), 033001 (2009).
27. C. A. Mack, "Stochastic modeling of photoresist development in two and three dimensions," *J. Micro/Nanolithogr. MEMS MOEMS* **9**(4), 041202 (2010).
28. C. A. Mack, "Defining and measuring development rates for a stochastic resist: a simulation study," *J. Micro/Nanolithogr. MEMS MOEMS* **12**(3), 033006 (2013).
29. T. B. Michaelson et al., "The effects of chemical gradients and photoresist composition on lithographically generated line edge roughness," *Proc. SPIE* **5753**, 368 (2005).
30. J. J. Biafore et al., "Mechanistic simulation of line-edge roughness," *Proc. SPIE* **6519**, 65190Y (2007).
31. J. Jiang et al., "Impact of acid statistics on EUV local critical dimension uniformity," *Proc. SPIE* **10143**, 1014323 (2017).
32. C. A. Mack, "Line-edge roughness and the ultimate limits of lithography," *Proc. SPIE* **7639**, 763931 (2010).
33. C. A. Mack, "Understanding the efficacy of linewidth roughness post-processing," *J. Micro/Nanolithogr. MEMS MOEMS* **14**(3), 033503 (2015).
34. C. Cutler et al., "Roughness power spectral density as a function of resist parameters and its impact through process," *Proc. SPIE* **10587**, 1058707 (2018).
35. V. Rutigliani et al., "Setting up a proper power spectral density (PSD) and autocorrelation analysis for material and process characterization," *Proc. SPIE* **10585**, 105851K (2018).
36. P. Naulleau et al., "Ultrahigh efficiency EUV contact-hole printing with chromeless phase shift mask," *Proc. SPIE* **9984**, 99840P (2016).

Chris A. Mack developed the lithography simulator PROLITH and founded the company FINLE Technologies in 1990. He received his SEMI Award for North America in 2003 and the SPIE Frits Zernike Award for Microlithography in 2009. He is a fellow of SPIE and IEEE and an adjunct faculty member at the University of Texas at Austin. In 2017 he cofounded Fractilia, where he now works as chief technical officer.