Systematic errors in the measurement of power spectral density

Chris A. Mack

Systematic errors in the measurement of power spectral density

Chris A. Mack
Lithoguru.com
1605 Watchhill Road
Austin, Texas 78703
E-mail: chris@lithoguru.com

Abstract. Measurement of the power spectral density (PSD) of a rough surface or a feature involves large random and systematic errors. While random errors can be reduced by averaging together many PSDs, systematic errors can be reduced only by carefully studying and understanding the sources of these systematic biases. Using both analytical expressions and numerical simulations for the measurement of the PSD of line-edge roughness, four sources of systematic errors are evaluated: aliasing, leakage, averaging, and image noise. Exact and approximate expressions for each of these terms are derived over a range of roughness exponents, allowing a measured PSD to be corrected for its systematic biases. In the absence of image noise, the smallest measurement bias is obtained when appropriate data windowing is used and when the sampling distance is set to twice the measurement signal width. Uncorrected PSD measurements are likely to systematically bias each of the PSD parameters, with the roughness exponent especially susceptible to bias. © *2013 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMM.12.3.033016]

1 Introduction

Line-edge roughness (LER) and line-width roughness (LWR) in lithography are best characterized by the power spectral density (PSD) of the roughness or similar measures of roughness frequency and correlation. In any real measurement, however, an approximation to the actual PSD is made by sampling the edge position (in the case of LER) or the line-width (in the case of LWR) of a finite-length feature. The result is called the discrete PSD, and it exhibits not only random errors (measuring noise is fundamentally noisy), but also systematic biases. Thus, it is important to understand the nature and the magnitude of these systematic errors in PSD measurement and to develop methods for their mitigation. While most studies of LER measurement bias have focused on the LER standard deviation,[1,2] this work will address biases in the PSD itself.

There are several tools available to study the biases in the PSD measurement. For the special case of a roughness exponent of 0.5, an analytical expression for the discrete PSD has been derived by Hiraiwa and Nishida.[3–7] For other cases, numerically generated synthetic rough edges, which are then "measured" and analyzed, leads to further insights into errors in the PSD measurement.[8,9] In this article, the properties of the measured PSD will be examined using these and other analytical tools with the goal of defining and then minimizing the systematic errors present in the PSD measurement.

2 Theory of the Discrete PSD

Given a randomly rough lithographic feature, such as a long line, different points along the edge of that feature may be correlated. To examine such correlations, the autocovariance function ($\tilde{R}$) of the feature edge position (or feature width for the case of two feature edges that are completely uncorrelated) is defined as

$$\tilde{R}(s,t) = \langle (w(s) - \bar{w})(w(t) - \bar{w}) \rangle, \tag{1}$$

where w is the measured line-width/edge position, s and t are the positions where measurements are made along the length of the line, $\bar{w}$ is the mean line-width/edge position of the feature, and $\langle \ldots \rangle$ represents the average over many instances of the roughness. If the process is stationary, the resulting autocovariance will be a function of only the distance $s - t$.

The LWR/LER PSD is generally calculated as the squared magnitude of the Fourier transform of the feature width/edge position. For any real measurement, though, the feature will be sampled typically with measurements made some fixed distance apart, Δy. The discrete PSD (PSD$_d$) will then be calculated from a discrete Fourier transform (such as the fast Fourier transform) of this data.

$$
\begin{aligned}
\langle \text{PSD}_d(f) \rangle &= \frac{\Delta y}{N} \left\langle \left| \sum_{s=0}^{N-1} (w(s) - \bar{w}) e^{-i2\pi\tau s/N} \right|^2 \right\rangle \\
&= \frac{\Delta y}{N} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} \tilde{R}(s,t) e^{i2\pi\tau(s-t)/N} \\
&= \Delta y \sum_{m=-(N-1)}^{N-1} \left(1 - \frac{|m|}{N}\right) \tilde{R}(m) e^{i2\pi\tau m/N}, \tag{2}
\end{aligned}
$$

where N is the number of measurement points, $L = \Delta y N$ is the length of the line being sampled, and the frequency $f = \tau/L$. The far right side of Eq. (2) shows that calculating

the PSD from the discrete Fourier transform of line-width data is equivalent to the discrete Fourier transform of the (biased) estimator for the autocovariance function (as expected from the Wiener–Khinchin theorem).

As Hiraiwa and Nishida have shown,[3] it is possible to calculate the discrete PSD analytically, given a certain model form for the autocovariance function. For example, it has been common to assume that a stretched exponential autocovariance function can be applied to rough features.

$$\tilde{R}(s-t) = \sigma^2 e^{-(|s-t|/\xi)^{2\alpha}}, \tag{3}$$

where ξ is the correlation length and α is the roughness exponent. For $\alpha = 0.5$, the resulting continuous PSD can be analytically derived.[10] For a one-dimensional problem (such as LER or LWR),

$$\text{PSD}(f) = \frac{2\sigma^2\xi}{1 + (2\pi f\xi)^2}. \tag{4}$$

Using the stretched exponential model for autocovariance and $\alpha = 0.5$, it is also possible to calculate the discrete PSD analytically. The result derived by Hiraiwa and Nishida (using slightly different notation here) is

$$\frac{\langle \text{PSD}_{\text{d}}(f)\rangle}{2\sigma^2\xi} = \delta\left(\frac{1}{2} + \text{Re}\left\{\frac{z}{1-z}\right\} - \frac{1}{N}\text{Re}\left\{\frac{z-z^{N+1}}{(1-z)^2}\right\}\right), \tag{5}$$

where $\delta = \Delta y/\xi$, $a = 2\pi f\xi$, and $z = e^{-\delta}e^{ia\delta}$.

It is important to note that this result is only valid for the case of a roughness exponent $\alpha = 0.5$. Experimental LER data often shows roughness exponents in the range of 0.7 to 0.8.[11] Unfortunately, an analytical solution for the discrete PSD is not possible for these cases. Other techniques for dealing with these higher roughness exponents will be described in later sections.

3 Properties of the Discrete PSD

In practice, Eq. (5) is awkward to deal with and is evaluated numerically. Some simplifications, however, will make the analytical PSD$_{\text{d}}$ more convenient. Taking each term separately

$$\delta\left(\frac{1}{2} + \text{Re}\left\{\frac{z}{1-z}\right\}\right) = \frac{\delta}{2}\text{Re}\left\{\frac{1+z}{1-z}\right\}$$
$$= \frac{\delta \sinh(\delta)}{2[\cosh(\delta) - \cos(a\delta)]}, \tag{6}$$

where this expression is used for $0 \le a\delta \le \pi$ (i.e., for frequencies at or below the Nyquist frequency). The second term in Eq. (5) can likewise be modified by making the reasonable assumption that $L \gg \xi$ (i.e., the length of line being measured is much larger than the correlation length, a requirement for the accurate PSD measurement). In this case, $|z^N| = e^{-L/\xi} \ll 1$ and

$$\frac{\delta}{N}\text{Re}\left\{\frac{z-z^{N+1}}{(1-z)^2}\right\} \approx \frac{\xi}{L}\delta^2\text{Re}\left\{\frac{z}{(1-z)^2}\right\}$$
$$= \frac{\xi}{L}\delta^2 \frac{\cosh(\delta)\cos(a\delta) - 1}{2[\cosh(\delta) - \cos(a\delta)]^2}. \tag{7}$$

Thus, the discrete PSD becomes

$$\frac{\langle \text{PSD}_{\text{d}}(f)\rangle}{2\sigma^2\xi} \approx \frac{\delta \sinh(\delta)}{2[\cosh(\delta) - \cos(a\delta)]}$$
$$\times \left(1 - \frac{\xi}{L}\frac{\delta}{\sinh(\delta)}\frac{\cosh(\delta)\cos(a\delta) - 1}{[\cosh(\delta) - \cos(a\delta)]}\right). \tag{8}$$

Further, Eq. (8) can be simplified by expanding the hyperbolic functions as Taylor series for the reasonable case of small δ (i.e., where the sampling distance is much smaller than the correlation length, also a requirement for the accurate measurement of the PSD). The resulting equation is quite simple.

$$\langle \text{PSD}_{\text{d}}(f)\rangle = \text{PSD}_{\text{c}}(f)(1 + \varepsilon_{\text{alias}})(1 + \varepsilon_{\text{leakage}}), \tag{9}$$

where

$$\text{PSD}_{\text{c}}(f) = \frac{2\sigma^2\xi}{1 + (2\pi f\xi)^2}$$

$$\varepsilon_{\text{leakage}} = \left(\frac{\xi}{L}\right)\left(\frac{(2\pi f\xi)^2 - 1}{(2\pi f\xi)^2 + 1}\right) + O\left(\frac{\xi}{L}e^{-L/\xi}\right), \quad \text{and}$$

$$\varepsilon_{\text{alias}} \approx \left(\frac{\pi f\Delta y}{\sin(\pi f\Delta y)}\right)^2 - 1.$$

Thus, the discrete PSD is equal to the continuous PSD modified by two error terms, $\varepsilon_{\text{alias}}$ and $\varepsilon_{\text{leakage}}$.

Sampling means the resulting PSD contains frequency information only up to the Nyquist frequency, $f = 1/(2\Delta y)$. Since the actual feature being measured contains frequency information higher than the Nyquist frequency, the power from these higher frequencies is added to frequencies at and below the Nyquist frequency in the sampled PSD, a phenomenon called "aliasing." This distortion can be significant, as will be shown below, and is captured by the term $\varepsilon_{\text{alias}}$. The aliasing term $\varepsilon_{\text{alias}}$ is 0 at $f = 0$, rising to about $\pi^2/4 - 1$ at the Nyquist frequency.

One can see that the second error term $\varepsilon_{\text{leakage}}$ varies from a low of $-\xi/L$ at $f = 0$ to a high approaching ξ/L at the highest frequencies, passing through zero at a frequency corresponding to the correlation length. The term "leakage" refers to the impact of measuring within a finite window (i.e., a finite length of the line being measured), resulting in a localized spreading of frequency components. As Eq. (9) shows, this leakage takes power away from the low-frequency components [frequencies below $1/(2\pi\xi)$] and adds power to the higher-frequency components of the PSD$_{\text{d}}$ [frequencies above $1/(2\pi\xi)$], effectively producing a slight blurring of the PSD. Leakage is minimized by making the measurement length large compared with the correlation length.

Thus, there are two sources of difference between the continuous and the discrete PSDs: the first is due to the nonzero value of ξ/L (as captured in the $\varepsilon_{\text{leakage}}$ term) and the second is due to the nonzero value of Δy (as captured in the $\varepsilon_{\text{alias}}$

term). Let us assume, for the sake of this analysis, that the true PSD behavior of a rough feature is given by the autocorrelation function of Eq. (3) with $\alpha = 0.5$, so that the true PSD is the continuous PSD of Eq. (4). Under this assumption, differences between the discrete and the continuous PSD are the result of sampling (the nonzero Δy and finite L), an artifact of the measurement process.

The analysis above shows that leakage decreases the apparent value of the zero-frequency PSD by $1 - \xi/L$. While leakage also affects the high-frequency terms, its impact is dwarfed in this case by the effects of aliasing. Aliasing has no impact on the zero frequency, but grows to a very significant level (a multiplicative factor of about 2.5 for the $\alpha = 0.5$ case) at the Nyquist frequency. Figure 1 shows plots comparing the discrete PSD [Eq. (5)], the continuous PSD [Eq. (4)], and the approximation to the discrete PSD [Eq. (9)]. For typical parameter values, the difference between the approximate discrete PSD and the exact discrete PSD is very small: an error of about $(\Delta y/\xi)^2/10$ at low frequencies decreasing to about $(\Delta y/\xi)^2/50$ at high frequencies. The alias and leakage terms can be plotted individually as well, as shown in Fig. 2.

Equation (9) was derived for the single case of $\alpha = 0.5$. However, the form of Eq. (9), with its ε_{alias} and $\varepsilon_{leakage}$ terms, will be applicable to any roughness exponent. Thus, the next goal will be to find more general approaches for determining these two error terms for other roughness exponents.

4 Calculating Aliasing Using the Kirchner Method

Kirchner[12] developed a method for calculating the effects of aliasing on a measured PSD (leakage not included). When a random signal is undersampled (meaning there is information in the signal at frequencies higher than the Nyquist frequency), the apparent (measured) spectral power at some frequency f_0 will contain not only the true power of the continuous PSD, but also the power at the aliased frequencies $kf_s \pm f_0$, where $f_s = 1/\Delta y$ is the sampling frequency and k is any positive integer. For a real-valued signal, where the PSD will be symmetric about $f = 0$, the resulting discretely measured PSD (including only the effects of aliasing) will be[12]
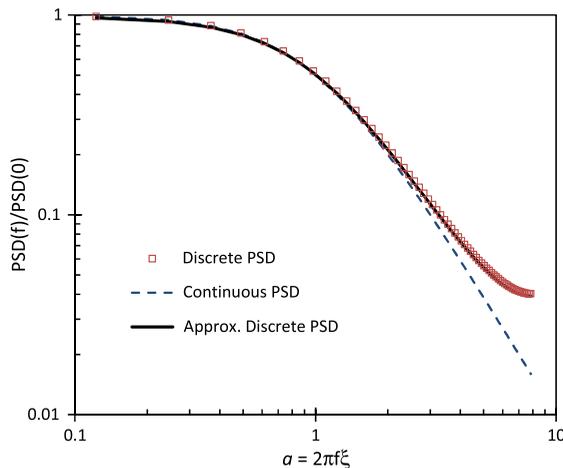
$$\langle \text{PSD}_d(f) \rangle = \text{PSD}(f) + \sum_{k=1}^{\infty}[\text{PSD}(kf_s - f)$$
$$+ \text{PSD}(kf_s + f)]. \tag{10}$$

Consider the PSD of Eq. (4) in the frequency range where $f \gg 1/(2\pi\xi)$, where we expect the aliasing to be the most significant. Kirchner's formula will then become

$$\langle \text{PSD}_d(f) \rangle \approx \frac{2\sigma^2\xi}{(2\pi f\xi)^2}$$
$$\times \left(1 + (f\Delta y)^2 \sum_{k=1}^{\infty}\left[\frac{1}{(k - f\Delta y)^2} + \frac{1}{(k + f\Delta y)^2}\right]\right). \tag{11}$$

The infinite summation converges to an analytical result, giving

$$\varepsilon_{alias} = (f\Delta y)^2 \sum_{k=1}^{\infty}\left[\frac{1}{(k - f\Delta y)^2} + \frac{1}{(k + f\Delta y)^2}\right]$$
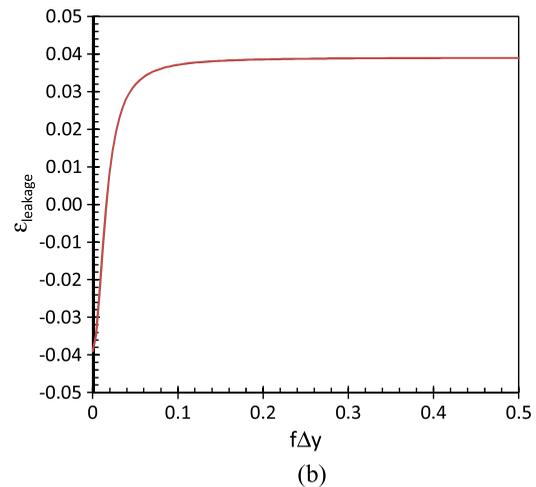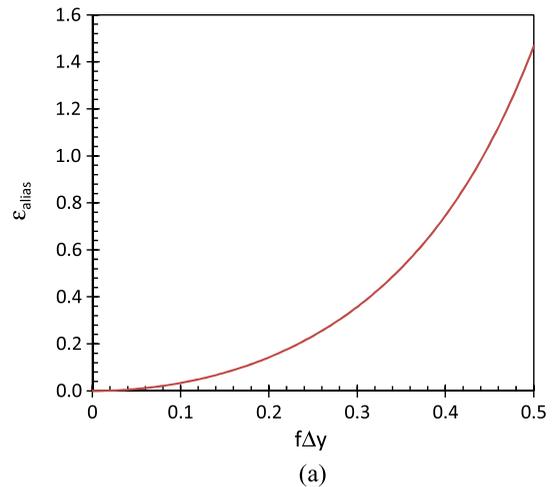$$= \frac{(\pi f\Delta y)^2}{\sin^2(\pi f\Delta y)} - 1. \tag{12}$$

(a)

(b)

Fig. 1 Plots of the discrete power spectral density (PSD) (symbols), the continuous PSD (dashed line), and the approximation to the discrete PSD (solid line) found in Eq. (9) using $N = 128$, $\xi = 6.4$ nm, and $\Delta y = 2.56$ nm.

Fig. 2 Plots of (a) the alias and (b) the leakage terms, ε_{alias} and $\varepsilon_{leakage}$, from Eq. (9) using $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm.

Thus, this discrete PSD including aliasing (but no leakage) matches the same result found in Eq. (9) for the case where $f \gg 1/(2\pi\xi)$.

While it is valuable to confirm using Kirchner's method the interpretation of $\varepsilon_{\text{alias}}$ in Eq. (9) as a term accounting for the effect of aliasing, its real value here is in its ability to numerically evaluate the impact of aliasing on the other PSD functions. Consider the Palasantzas extension of the PSD function of Eq. (4) to other roughness exponents:[13]

$$PSD(f) = \frac{PSD(0)}{[1 + (2\pi f \xi)^2]^{H+d/2}}, \quad (13)$$

where H plays the role of the Hurst (roughness) exponent, d is the dimensionality of the problem, and PSD(0) is adjusted to give the desired variance. For $d = 1$,

$$PSD(0) = 2\sigma^2 \xi \left(\frac{\sqrt{\pi}\Gamma\left(H + \frac{1}{2}\right)}{\Gamma(H)} \right). \quad (14)$$

This definition of the roughness exponent H matches the roughness exponent α defined by Eq. (3), when $H = \alpha = 0.5$, though not for other values. Again considering the frequency range where $f \gg 1/(2\pi\xi)$, Kirchner's formula will become

$$\langle PSD_d(f) \rangle \approx \frac{PSD(0)}{(2\pi f \xi)^{2H+1}}$$
$$\times \left(1 + (f\Delta y)^{2H+1} \sum_{k=1}^{\infty} \left[\frac{1}{(k - f\Delta y)^{2H+1}} + \frac{1}{(k + f\Delta y)^{2H+1}} \right] \right). \quad (15)$$

By analogy with our previous results, we can define the aliasing term as

$$\varepsilon_{\text{alias}} = (f\Delta y)^{2H+1} \sum_{k=1}^{\infty} \left[\frac{1}{(k - f\Delta y)^{2H+1}} + \frac{1}{(k + f\Delta y)^{2H+1}} \right]. \quad (16)$$

Since the summation will converge for $H > 0$, we can numerically evaluate $\varepsilon_{\text{alias}}(f)$ for different values of H. Some results are shown in Fig. 3, where carrying out the summation in Eq. (16) to $k = 100$ is sufficient.

The results of the numerical calculations of $\varepsilon_{\text{alias}}(f)$ for $0.5 \leq H \leq 1$ can be fit extremely well to a simple empirical expression. Let $\varepsilon_{0.5}$ be the analytical aliasing term for the case of $\alpha = 0.5$ [i.e., Eq. (12)],

$$\varepsilon_{\text{alias}}(f) \approx \left[1 - 0.421 \left(\frac{2H - 1}{H} \right) \right] (\varepsilon_{0.5})^{1+0.686(2H-1)}. \quad (17)$$

This empirical expression for $\varepsilon_{\text{alias}}$ produces a PSD that matches the one produced using Eq. (16) to better than 0.35% over the full range of frequency and Hurst exponents. While the numerical evaluation of Eq. (16) is simple and fast, Eq. (17) may prove useful when fitting to experimental data that is aliased.

5 Using Simulation to Determine Leakage and Aliasing

The Hiraiwa and Nishida discrete PSD function of Eq. (5), as simplified in Eq. (9), provides analytical equations for leakage and aliasing for the case of $\alpha = 0.5$. The Kirchner method provides a very simple numerical scheme for calculating the effects of aliasing for any PSD. For the Palasantzas PSD commonly used to model LER data, the Kirchner aliasing results are conveniently summarized in approximate form by Eq. (17) for values of roughness exponent H between 0.5 and 1.0. The only thing remaining is a determination of the leakage term for the roughness exponents other than 0.5.

Numerical simulation of rough features with predefined statistical properties provides a valuable numerical tool for determining the effects of leakage and aliasing, since leakage and aliasing can be individually turned on and off in such simulations. Here, the Thorsos method[14,15] was used to generate random rough edges with a Gaussian distribution and correlation behavior determined by the Palasantzas PSD. Letting L_s be the length of the simulated line, leakage will occur when the metrology length L is less than L_s. Letting Δs be the simulation grid size, aliasing will occur when the metrology sampling distance Δy is greater than Δs. Thus, leakage in the extraction of the PSD from measurement of the simulated line can be turned off by setting $L = L_s$. Likewise, aliasing in the simulation can be turned off by letting $\Delta y = \Delta s$.

As a first test, random rough lines were generated and their PSDs were extracted using $L = L_s$ and $\Delta y = \Delta s$. The PSDs of M simulations were averaged together, and the RMS relative difference between the resulting measured PSD and the input PSD was calculated. Such calculations were repeated multiple times to reduce the statistical uncertainty in the calculated RMS differences. As shown in Fig. 4, the measured PSD from the simulated rough lines has a relative uncertainty of 1.0 (for the case of $M = 1$), as expected. Averaging multiple PSDs together allows the measured PSD to converge to the input PSD as $1/\sqrt{M}$, also as expected. Note that this convergence trend shows that the measured PSD exhibits neither leakage nor aliasing (and thus, systematic differences between the discrete and the continuous
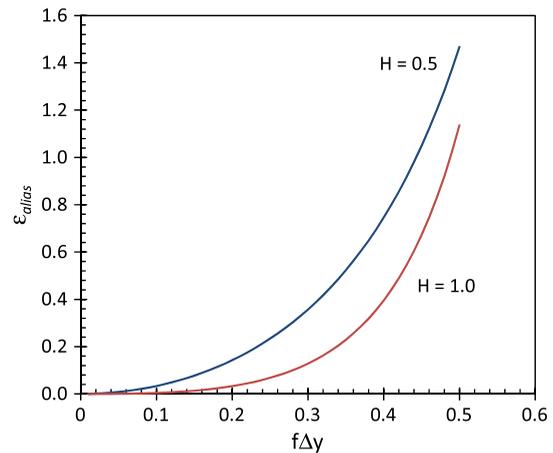

Fig. 3 Calculations of $\varepsilon_{\text{alias}}$ using Eq. (16) for two different roughness exponents. Results for $0.5 < H < 1.0$ produce aliasing between these two curves.
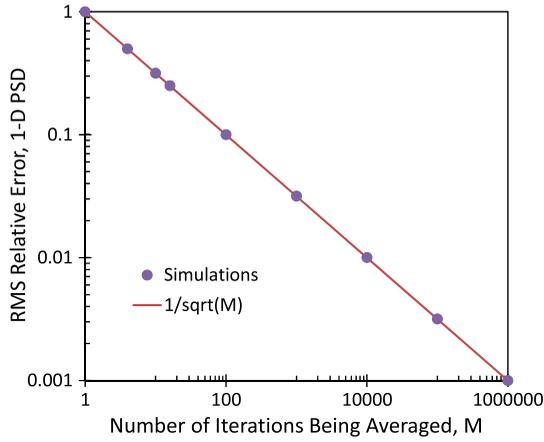
Fig. 4 Convergence of the numerically generated PSD (without leakage of aliasing) to the input PSD as a function of the number of trials being averaged together ($\sigma = 5$ nm, $\xi = 10$ nm, $\alpha = 0.5$, $\Delta y = 1$ nm, $N = 1024$). The standard $1/\sqrt{M}$ convergence trend is shown as the solid line with simulations shown as the symbols. The RMS relative error is the RMS relative difference between the measured PSD and the continuous PSD used as the input to the simulations (from Ref. 15).

PSDs) when $L = L_s$ and $\Delta y = \Delta s$. It is important to note that the RMS relative error in the PSD shown here has nothing to do with noise in the measurement of the line-edge position, but rather is a consequence of the randomness of the line samples being measured. As the sample size increases (more PSDs are averaged together), the result converges to the expected value.

Leakage can be turned on for the simulations without aliasing by letting $L_s = 2L$ and keeping $\Delta y = \Delta s$. From the simulations, an "experimental" leakage term can be calculated from

$$\langle \text{PSD}_{\text{simulation}}(f) \rangle = \text{PSD}(f)[1 + \varepsilon_{\text{leakage}}(1 + \varepsilon_{\text{alias}})]. \quad (18)$$

Note that the impact of aliasing on the leakage is still present in the simulations even though the main aliasing term is absent. Figure 5 compares the experimental leakage (as determined from the average of 400 million simulations) with the derived expression for $\varepsilon_{\text{leakage}}$ in Eq. (9) for the case of $\alpha = 0.5$. As can be seen, the simulations and the derived analytical expression match extremely well until the very highest frequencies. At the Nyquist frequency, the simulated $\varepsilon_{\text{leakage}}$ is higher by about 9%, resulting in a difference in the simulated and the predicted PSD of <0.9%. The reason for this difference at high frequencies is unclear, but it is small enough to be of little concern. Simulations using different sampling distances, line lengths, and correlation lengths produced similar results.

Likewise, leakage can be turned off by setting $L_s = L$, and aliasing can be turned on by setting $\Delta s = \Delta y/N_{\text{sim}}$. Comparison with the analytical aliasing term is, however, complicated by the fact that the analytical result assumes a continuum, i.e., $N_{\text{sim}} \to \infty$. To understand the impact of the simulation grid size on the aliasing results, N_{sim} was varied from 2 to 128, and the aliasing term $\varepsilon_{\text{alias}}$ was calculated from the measured discrete PSD. The results are shown in Fig. 6 (10 million simulations per curve). As N_{sim} increases, the aliasing converges to the continuum result from Eq. (9). The rate of convergence is well described by
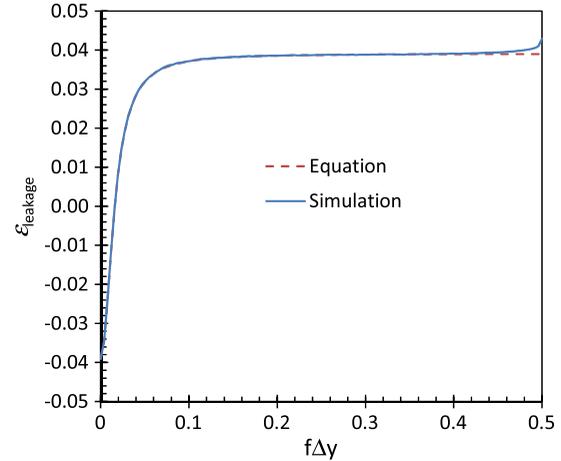

Fig. 5 Plots of $\varepsilon_{\text{leakage}}$ from Eq. (9) (dashed line) and from simulations (solid line) with aliasing turned off using $\alpha = 0.5$, $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm.

$$\varepsilon_{\text{alias}}(\text{simulation}) = \varepsilon_{\text{alias}}(\text{continuum}) - \frac{1}{N_{\text{sim}}}(2f\Delta y)^2. \quad (19)$$

Thus, the worst-case difference is at the Nyquist frequency ($2f\Delta y = 1$), where the simulation approaches the continuum answer with a difference equal to $1/N_{\text{sim}}$. Based on this result, the simulations below will use $N_{\text{sim}} = 128$, and the calculated $\varepsilon_{\text{alias}}$ will be corrected by adding $(2f\Delta y)^2/N_{\text{sim}}$ to give the best approximation to the continuum value of $\varepsilon_{\text{alias}}$.

Comparing these simulation results for $\varepsilon_{\text{alias}}$ with the Kirchner calculations, using a finite value of N_{sim} is equivalent to using a finite range of k in the summation in Eq. (16). In fact, the simulation results shown in Fig. 6 can be reproduced almost exactly using the Kirchner equation and letting the summation go to a maximum k of $N_{\text{sim}}/2$. Physically, the use of a maximum k in the Kirchner summation or a finite
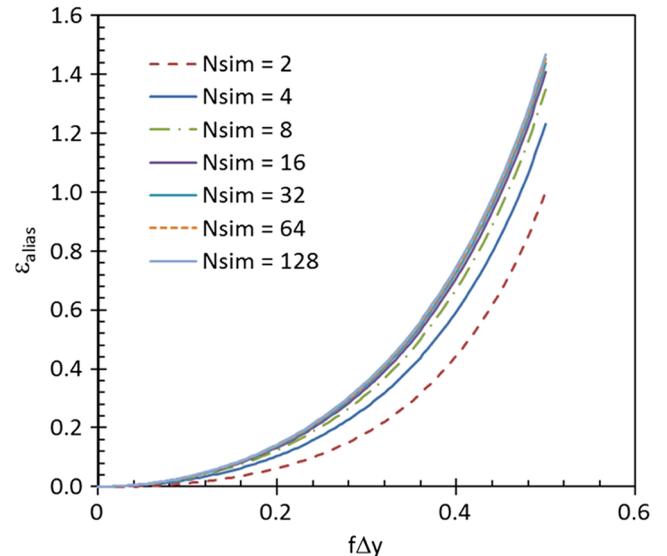

Fig. 6 Plots of $\varepsilon_{\text{alias}}$ from simulations as a function of the ratio of the metrology sampling distance to the simulation grid size (N_{sim}) using $\alpha = 0.5$, $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm.

N_{sim} in the simulations is equivalent to saying there is a maximum frequency present in the physical feature being measured, and that the real world ceases to be a continuum at a small enough length scale.

Figure 7 compares ε_{alias} from Eq. (9) with the results of eight million simulations. Both curves are plotted on the same graph, but the results are indistinguishable, with differences <0.002. The results shown in Figs. 5–7 for $\alpha = 0.5$ confirm that simulations are capable of elucidating the roles of leakage and aliasing on the resulting PSD with great accuracy. These simulations have also confirmed the accuracy of the analytical results represented in Eq. (9). This same simulation approach can now be used to determine our one unknown factor: how does the leakage term change as a function of roughness exponent? The simulated impact of the roughness exponent H on the leakage term is shown in Fig. 8. While the impact of leakage is small for the case of $H = 0.5$, it is much larger for larger roughness exponents.

6 Reducing Leakage with Data Windowing

The above sections described various tools for calculating the amount of aliasing and leakage in the measurement of PSD. The impact of these biases on the measurement can be reduced in two ways: numerically correcting the measured PSD for aliasing and leakage or designing a measurement process that has inherently small leakage and aliasing. One common way to reduce leakage is by using data windowing. In this approach, the measurement value $w(s)$ used in Eq. (2) is weighted by a window $g(s)$ before taking the discrete Fourier transform. Standard LER measurement can be thought of as applying a rectangular measurement window to a long feature: in the region of the line being measured $g(s) = 1$ and outside the region of line being measured $g(s) = 0$. Note that the convolution of this rectangular window with itself produces the $(1 - |m|/N)$ term in Eq. (2) that biases the estimator for the autocovariance.

The impact of the data window on the PSD can be seen by considering a continuous measurement of the PSD over a finite line length:

$$\langle PSD_{measure}(f)\rangle = \left\langle \left| \int_{-\infty}^{\infty} g(y)(w(y) - \bar{w})e^{-i2\pi fy}dy \right|^2 \right\rangle$$
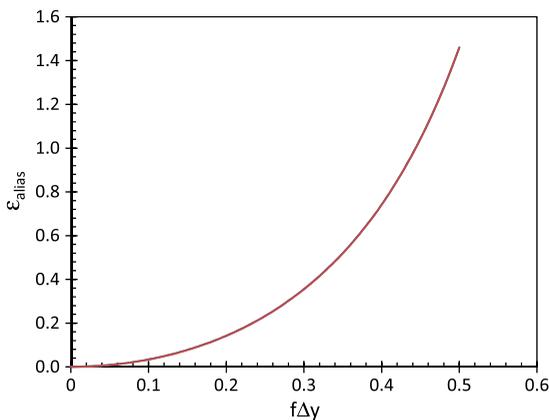$$= G^2(f) \otimes PSD(f), \qquad (20)$$

Fig. 7 Plots of ε_{alias} from Eq. (9) and from simulations with leakage turned off using $\alpha = 0.5$, $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm. The two curves are indistinguishable.

where $g(y)$ is assumed to be symmetric about $y = 0$, so that $G(f)$, the Fourier transform of $g(y)$, will be real. Measuring the LER using a data window $g(y)$ results in a measured PSD that is equal to the continuous PSD convolved with the square of $G(f)$. For the rectangular window of a conventional LER measurement, the continuous PSD is convolved with

$$G^2(f) = \left(\frac{\sin(\pi f L)}{\pi f L} \right)^2. \qquad (21)$$

As L becomes large, window term of Eq. (21) approaches a delta function, and the measured PSD becomes a perfect reproduction of the continuous PSD. For finite L, the convolution of the window term causes a "leakage" of other frequencies into the measured PSD at f.

Note that Eq. (21) falls off as $1/f^2$ away from the frequency being measured. The PSD, on the other hand, falls off as $1/f^{2H+1}$ at high frequencies. For $H = 0.5$, the fall-off of the window convolution term exactly matches the rise of the PSD toward lower frequencies, so that the amount of leakage is a constant at high frequencies. For $H > 0.5$, the PSD rises faster than the fall-off of the window convolution term, and the leakage term gets bigger for higher frequencies. Thus, leakage can be reduced for $0.5 < H < 1.0$ by using a $G^2(f)$ that falls off faster than $1/f^3$. There are a number of data windows commonly employed in signal analysis that exhibit this property.

Consider the Bartlett window[16] given by

$$g_{Bartlett}(y) = \begin{cases} 2 - 4|y|/L, & -L/2 < y < L/2 \\ 0, & \text{otherwise} \end{cases}. \qquad (22)$$

The Bartlett window is just an isosceles triangle with base width of L and height adjusted to give the same area as the rectangular window. The Fourier transform of the Bartlett window gives

$$G^2_{Bartlett}(f) = \left(\frac{\sin(\pi f L/2)}{\pi f L/2} \right)^4. \qquad (23)$$

Since this window term falls off as $1/f^4$, the high frequencies of the PSD will not experience significant leakage. Other common windows, such as the Welch and Hann windows, have the same behavior.[16] Figure 9 shows simulations of the measured PSD using the Bartlett window, extracting the leakage term as before. Note that the resulting leakage is <2% for all frequencies, and is thus small enough to be ignored under most circumstances. The small rise in leakage at the Nyquist frequency matches the difference seen between simulated and analytical leakage terms shown in Fig. 5.

Leakage can also be estimated by considering Eq. (2) in the continuum limit of $\Delta y = 0$. The PSD with leakage (but no aliasing) becomes
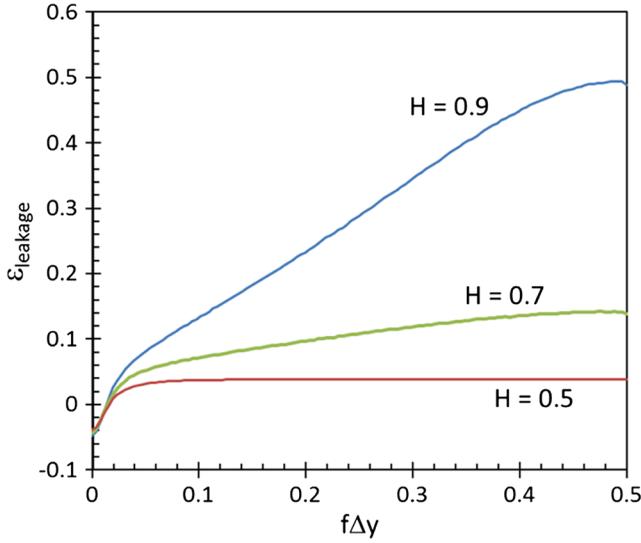
Fig. 8 Plots of $\varepsilon_{\text{leakage}}$ from simulations for different values of the roughness exponent H ($N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm).

$$\langle \text{PSD}_{\text{measure}}(f) \rangle = \int_{-\infty}^{\infty} [g(y) \otimes g(y)] \tilde{R}(y) e^{-i2\pi f y} dy$$

$$= 2 \int_{0}^{L} [g(y) \otimes g(y)] \tilde{R}(y) \cos(2\pi f y) dy. \quad (24)$$

For the case of a rectangular data window,

$$g(y) \otimes g(y) = \begin{cases} 1 - |y|/L, & -L < y < L \\ 0, & \text{otherwise} \end{cases}. \quad (25)$$

When $\alpha = 0.5$, plugging Eq. (25) into Eq. (24) produces Eq. (9) without the aliasing term. For the Bartlett and the other "higher order" data windows, the convolution of the window with itself produces a y^2 term rather than the linear term of Eq. (25). For $\alpha = 0.5$, this produces

$$\varepsilon_{\text{leakage}} = 12 \left(\frac{\xi}{L} \right)^2 \left(\frac{3(2\pi f \xi)^2 - 1}{[(2\pi f \xi)^2 + 1]^2} \right). \quad (26)$$

Equation (26) matches the simulation results quite well. For other values of the roughness exponent H, the results of Fig. 9 can be approximated by increasing the leakage magnitude:

$$\varepsilon_{\text{leakage}} \approx 24H \left(\frac{\xi}{L} \right)^2 \left(\frac{3(2\pi f \xi)^2 - 1}{[(2\pi f \xi)^2 + 1]^2} \right). \quad (27)$$

7 Impact of Spatial Averaging

One useful approach for reducing the effects of aliasing is through averaging. If the spacing between measurements is Δy, the measurement can be (and usually is) the average line-width or -edge position over some range η. If $\eta = 0$, we have the measurement at a point, as was assumed above in the derivation of the discrete PSD and in the simulations. For $\eta > 0$, the averaging dampens the high-frequency components of the signal, and thus, the aliasing. The impact of this kind of averaging has been previously derived[12] with

the PSD including averaging equal to the PSD assuming no averaging multiplied by the square of the Fourier transform of the averaging shape function. For a simple rectangular shape (straight averaging over the distance η), the Fourier transform is a sinc function giving

$$\text{PSD}_{\text{d-avg}}(f) = \text{PSD}_{\text{d}}(f) \left(\frac{\sin(\pi f \eta)}{\pi f \eta} \right)^2. \quad (28)$$

Consider the case of $H = 0.5$. Since the alias term is, in fact, a sinc function, choosing $\eta = \Delta y$ above gives the product of aliasing and averaging $= 1$ for all frequencies. In other words, proper averaging can greatly reduce (and theoretically even eliminate) aliasing.

When measuring LER using a scanning electron microscope (SEM), the measurement spot can be assumed to be a Gaussian. A Gaussian-shaped beam of electrons interacts with the feature being measured to produce a Gaussian-shaped measurement signal (wider than the incident beam) of full-width-half-maximum (FWHM) η. The impact of this averaging can be seen in Fig. 10 using simulation and is a function of $\eta/\Delta y$. For no averaging ($\eta = 0$), aliasing makes the measured PSD higher at the high frequencies. Averaging lowers the measured PSD at high frequencies, thus reducing the impact of aliasing. However, for $\eta > \Delta y/2$, the impact of averaging is greater than aliasing, and the measured PSD is suppressed at high frequencies.

Consider typical SEM measurement of LER. A typical SEM incident spot size is about 2 nm. As this spot interacts with the material being measured, scattered electrons within the material grow the interaction volume, so that the measurement signal would typically be 4 to 6 nm wide depending on the electron energy and materials involved. If the sampling distance is set to 4 nm (a commonly recommended value), then averaging would occur over a distance of 1 to 1.5 Δy. As shown in Fig. 10, the result will be a PSD with suppressed high-frequency power and will appear to have a higher value of the roughness exponent. Note that the
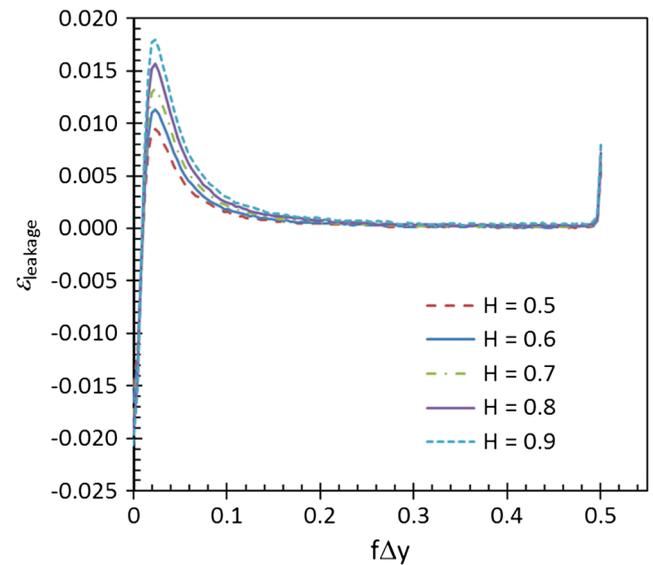

Fig. 9 Plots of $\varepsilon_{\text{leakage}}$ from simulations using the Bartlett window for different values of the roughness exponent H ($N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm).

measurement by atomic force microscopy will also produce averaging with η about equal to the probe diameter.

The Kirchner method can be accommodated to calculate the effects of averaging. As was shown in Eq. (28), the PSD including both aliasing and averaging is equal to the PSD with aliasing multiplied by the square of the Fourier transform of the measurement signal shape. For the case of a Gaussian signal of FWHM η,

$$\text{PSD}_{d-avg}(f) = \text{PSD}_d(f)e^{-(2\pi\sigma_m f)^2}$$
$$= \text{PSD}_d(f)\exp\left(-\frac{(\pi\eta f)^2}{2\ln(2)}\right), \tag{29}$$

where σ_m is the standard deviation of the Gaussian measurement signal. This expression was derived based on the use of the continuum Fourier transfer and ignoring leakage and produces a result that matches simulation when $\eta \gg \Delta y$ and $\xi \ll L$.

Averaging can also be accounted for by including its effects into the autocovariance function. Taking the inverse Fourier transform of Eq. (29) produces the "averaged" autocovariance function:

$$\tilde{R}_{avg}(x) = \tilde{R}(x) \otimes \left(\frac{1}{2\sqrt{\pi}\sigma_m}e^{-x^2/4\sigma_m^2}\right). \tag{30}$$

For the case of $\alpha = 0.5$, this gives

$$\tilde{R}_{avg}(x) = \frac{\sigma^2}{2}e^{(\sigma_m/\xi)^2}\left[e^{-x/\xi}\text{erfc}\left(\frac{\sigma_m}{\xi} - \frac{x}{2\sigma_m}\right)\right.$$
$$\left. + e^{x/\xi}\text{erfc}\left(\frac{\sigma_m}{\xi} + \frac{x}{2\sigma_m}\right)\right]. \tag{31}$$

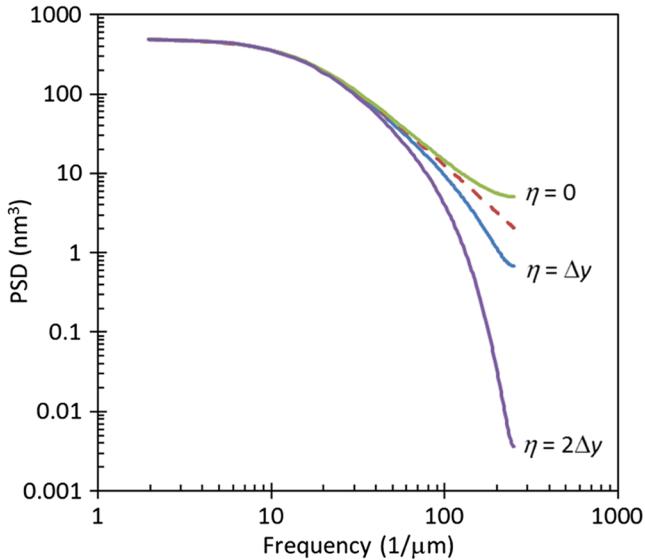The averaged autocorrelation function is plotted in Fig. 11.

Fig. 10 Simulations of the impact of averaging on the measured PSD ($N = 256$, $\xi = 10$ nm, $H = 0.5$, and $\Delta y = 2$ nm, and rectangular measurement window). The full-width-half-maximum (FWHM) of the Gaussian measurement signal (η) is varied from 0 to twice the sampling distance. The continuous PSD (without aliasing, leakage, or averaging) is shown as the dotted line.

For the important case where one cannot assume the Gaussian averaging width to be much greater than the sampling grid size, numerical simulations can be used to understand the impact of averaging. As shown in Fig. 10 (for the case of $H = 0.5$), a Gaussian FHWM of about $\Delta y/2$ produces an averaging effect that cancels the aliasing over most of the frequency range. Further simulations for other roughness exponents confirmed this simple rule of thumb. Figure 12 shows the relative difference between the discrete PSD, including aliasing, leakage, and averaging, and the continuous PSD for $H = 0.5$ and 0.9. With averaging near its optimum ($\eta = \Delta y/2$), the discrete PSD differs from the continuous PSD by <5% out to 75% of the Nyquist frequency. When using the Welch data window rather than a rectangular window, setting $\eta = \Delta y/2$ produces a discrete PSD that differs from the continuous PSD by <11% out to 90% of the Nyquist frequency for H between 0.5 and 0.9.

8 Impact of SEM Image Noise

Statistical noise in the grayscale image produced by the SEM has been shown to cause a positive bias in the LER σ.[1,17] The image noise, a random variation in the pixel grayscale values within the image caused by electron shot noise and other sources of variation, results in measurement noise in the edge position/line-width $w(s)$ and is generally thought to be white noise, at least over the frequency range typical of PSD measurement. Thus, a standard model for SEM image noise is[17]

$$\text{PSD}_{w/noise}(f) = \text{PSD}_{w/o\ noise}(f) + \sigma_{noise}^2\Delta y, \tag{32}$$

where σ_{noise} is the SEM image noise contribution, the standard deviation of the edge position/line-width $w(s)$ caused by pixel grayscale variations. Integrating this expression over all frequencies, up to the Nyquist frequency, produces the well-known biased LER result:[1]

$$\sigma_{w/noise}^2 = \sigma_{w/o\ noise}^2 + \sigma_{noise}^2. \tag{33}$$

SEM measurement noise can be added to the simulations by adding an uncorrelated normally distributed random noise
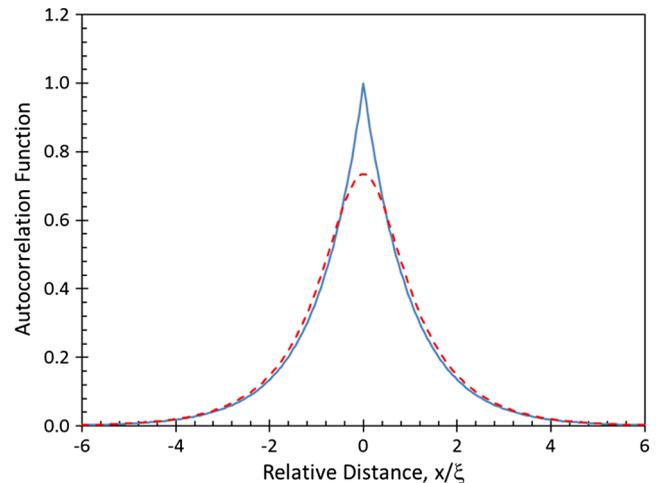

Fig. 11 Plots of the standard exponential autocorrelation function (solid line) and the Gaussian-averaged autocorrelation function (dashed line) for $\alpha = 0.5$ and $\sigma_m = 0.3\xi$.
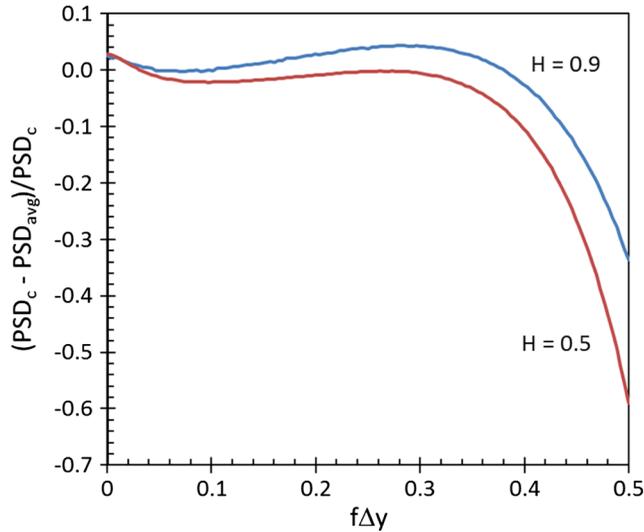
Fig. 12 The relative difference between the averaged PSD (including aliasing and leakage) and the continuous PSD (assumed to be the true value) when $\eta = \Delta y/2$ ($N = 256$, $\xi = 10$ nm, $\Delta y = 2$ nm, $H = 0.5$ and 0.9, and rectangular measurement window). Each curve is the average of more than 1 million simulations.

value to each simulated line-width/edge position. Figure 13 shows the impact of increasing SEM noise on the resulting PSD. As can be seen, this white noise has little impact on the low-frequency PSD values, but can dramatically increase the high-frequency PSD. Subtracting each of the PSDs in Fig. 13 from the no-noise PSD produces a constant value that follows Eq. (32) almost exactly.

Considering σ_{noise} as the fourth parameter in the PSD model, its value can be estimated in the same way as σ, ξ, and H. Several methods for producing a bias-free measurement of LER (i.e., to eliminate the impact of σ_{noise}) have been proposed,[1,18–20] but less work has been done to assess the impact of these approaches on the measurement of PSD and to extract the roughness exponent and correlation
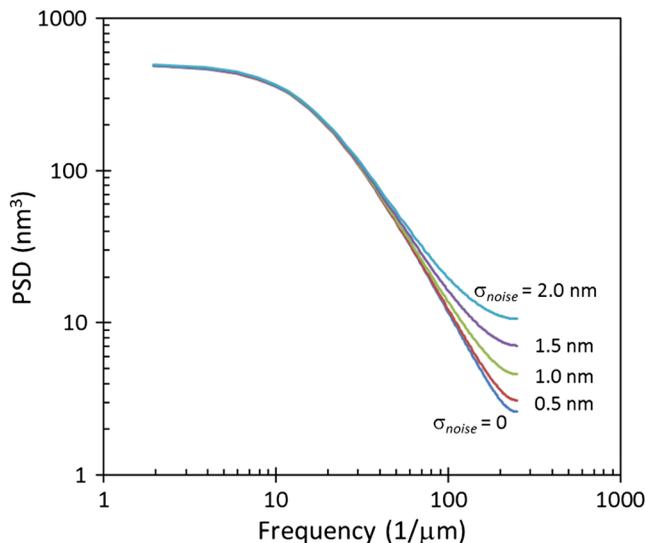

Fig. 13 Simulated impact of scanning electron microscope (SEM) measurement noise on the extracted PSD ($N = 256$, $\xi = 10$ nm, $\Delta y = 2$ nm, $H = 0.5$, $\eta = 1$ nm, and Welch measurement window). Each curve is the average of at least 10 million simulations.

length.[21,22] Note that pixel averaging along the length of the line, sometimes used to combat SEM image noise, will reduce high-frequency LER, as discussed in the previous section, and could possibly counter the increase in the high-frequency PSD exhibited in Fig. 13.

9 Conclusions

Systematic errors in PSD measurement are caused by several factors. Spectral leakage results from the finite value of L/ξ, the ratio of the measurement length to the correlation length. Aliasing occurs when the object being measured has power at frequencies greater than the sampling frequency. Averaging occurs whenever the measurement spot size is an appreciable fraction of the sampling distance. SEM image noise increases the PSD for all frequencies. All of these systematic errors can be significant and vary in degree and form as a function of the physical parameters of the PSD, in particular the correlation length and the roughness exponent. Note that each of these biases in PSD measurement is an artifact of the measurement process, and thus hides the true roughness behavior of the feature being measured. A thorough understanding of these effects can be used to minimize and/or correct the systematic errors, resulting in a measured PSD much closer to the actual PSD.

Several tools have been used to understand PSD measurement. The Hiraiwa and Nishida equation[3] gives an analytical, exact expression for the measured PSD, including aliasing and leakage (but not averaging), for the case of $H = 0.5$. An approximate form of the Hiraiwa and Nishida equation was derived in this article to explicitly show the separate effects of aliasing and leakage. The Kirchner equation[12] allows a simple numerical calculation of aliasing for any value of the roughness exponent. This equation can be extended to include the impact of averaging under limited conditions. Finally, simulation has been used to generate and measure random rough edges and to extract the various error terms. All three methods produce essentially identical results in the areas where their domains overlap.

Through the use of these numerical and analytical tools, a thorough understanding of many of the systematic biases in PSD measurement has been presented. Further, several mitigation strategies have been explored to reduce the error in the PSD measurement. The basic lessons learned are:

- Average together as many PSDs as possible to reduce random errors (100 averaged PSDs results in 10% random error in the PSD).

- Leakage scales as ξ/L and aliasing scales as $\Delta y = L/N$, so that a large N (the number of measurement points) is beneficial for both low leakage and low aliasing at the important mid frequencies (which is equivalent to requiring that $\Delta y \ll \xi \ll L$).

- Use data windowing (using the Bartlett, Welch, or similar window) to reduce spectral leakage to negligible levels.

- Balance aliasing with averaging by optimizing the sampling distance with respect to the spot size of the measurement signal. In the absence of SEM image noise, an optimum balance occurs when the sampling distance is set to about twice the spot size FWHM.

- Reduce SEM measurement noise as much as possible or eliminate it using a bias-free measurement scheme.

If this is not practical, extract the measurement noise from the measured PSD itself.

- Extract and report all three PSD parameters (σ, ξ, and H) as all three numbers are essential for understanding LER. The systematic biases in PSD measurement make accurate measurement of the roughness exponent especially difficult, and care should be taken in extracting and reporting this number.

The use of these strategies can result in a measured PSD close to the actual PSD. The use of measurement conditions significantly different from the optimum can result in a measured PSD far different from the actual PSD, especially at high frequencies. The SEMI standard for LER/LWR measurement[23] may or may not provide reasonable results depending on the value of the correlation length.

Correcting the measured PSD for systematic biases is possible using the techniques developed above, but requires that η, the measurement signal FWHM, to be known. Measuring the PSD with a SEM that has an unknown value of η produces a PSD with unknown biases. Further work on this topic should also include a method for evaluating the uncertainty in the values of σ, ξ, and H extracted from a measured PSD as a function of measurement parameters.

References

1. J. S. Villarrubia and B. D. Bunday, "Unbiased estimation of linewidth roughness," *Proc. SPIE* **5752**, 480–488 (2005).
2. A. Yamaguchi et al., "Bias-free measurement of LER/LWR with low damage of CD-SEM," *Proc. SPIE* **6152**, 61522D (2006).
3. A. Hiraiwa and A. Nishida, "Discrete power spectrum of line width roughness," *J. Appl. Phys.* **106**(7), 074905 (2009).
4. A. Hiraiwa and A. Nishida, "Spectral analysis of line edge and line-width roughness with long-range correlation," *J. Appl. Phys.* **108**(3), 034908 (2010).
5. A. Hiraiwa and A. Nishida, "Statistical- and image-noise effects on experimental spectrum of line-edge and line-width roughness," *J. Micro/Nanolith. MEMS MOEMS* **9**(4), 041210 (2010).
6. A. Hiraiwa and A. Nishida, "Image-pixel averaging for accurate analysis of line-edge and linewidth roughness," *J. Micro/Nanolith. MEMS MOEMS* **10**(2), 023010 (2011).
7. A. Hiraiwa and A. Nishida, "Statistical-noise effect on power spectrum of long-range-correlated line-edge and line-width roughness," *J. Micro/Nanolith. MEMS MOEMS* **10**(3), 033008 (2011).
8. P. Naulleau and J. Cain, "Experimental and model-based study of the robustness of line-edge roughness metric extraction in the presence of noise," *J. Vac. Sci. Technol.* **B25**(5), 1647–1657 (2007).
9. V. Constantoudis et al., "Line edge roughness and critical dimension variation: fractal characterization and comparison using model functions," *J. Vac. Sci. Technol.* **B22**(4), 1974–1981 (2004).
10. C. A. Mack, "Analytic form for the power spectral density in one, two, and three dimensions," *J. Micro/Nanolith. MEMS MOEMS* **10**(4), 040501 (2011).
11. A. R. Pawloski et al., "Line edge roughness and intrinsic bias for two methacrylate polymer resist systems," *J. Microlith. Microfab. Microsyst.* **5**(2), 023001 (2006).
12. J. W. Kirchner, "Aliasing in $1/f^\alpha$ noise spectra: origins, consequences, and remedies," *Phys. Rev. E* **71**(6), 061100 (2005).
13. 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).
14. E.I. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," *J. Acoust. Soc. Am.* **83**(1), 78–92 (1988).
15. C. A. Mack, "Generating random rough edges, surfaces, and volumes," *Appl. Opt.* **52**(7), 1487–1496 (2013).
16. W. H. Press et al., *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed., pp. 545–549, Cambridge University Press, Cambridge (1992).
17. A. Hiraiwa and A. Nishida, "Statistical- and image-noise effects on experimental spectrum of line-edge and line-width roughness," *J. Micro/Nanolith. MEMS MOEMS* **9**(4), 041210 (2010).
18. R. Katz et al., "Bias reduction in roughness measurement through SEM noise removal," *Proc. SPIE* **6152**, 61524L (2006).
19. A. Yamaguchi et al., "Single-shot method for bias-free LER/LWR evaluation with little damage," *Microelectron. Eng.* **84**(5–8), 1779–1782 (2007).
20. S.-B. Wang et al., "Practical and bias-free LWR measurement by CDSEM," *Proc. SPIE* **6922**, 692222 (2008).
21. V. Constantoudis and E. Gogolides, "Noise-free estimation of spatial line edge/width roughness parameters," *Proc. SPIE* **7272**, 72724B (2009).
22. V. Constantoudis, V.-K. M. Kuppuswamy, and E. Gogolides, "Effects of image noise on contact edge roughness and critical dimension uniformity measurement in synthesized scanning electron microscope images," *J. Micro/Nanolith. MEMS MOEMS* **12**(1), 013005 (2013).
23. SEMI Standard P47-0307, "Test method for evaluation of line-edge roughness and linewidth roughness," SEMI, San Jose, CA (2006).

Chris A. Mack developed the lithography simulation software PROLITH and founded and ran the company FINLE Technologies for 10 years. He then served as vice president of Lithography Technology for KLA-Tencor for 5 years, until 2005. In 2003, he received the SEMI Award for North America for his efforts in lithography simulation and education, and in 2009, he received the SPIE Frits Zernike Award for Microlithography. He is also an adjunct faculty member at the University of Texas at Austin. Currently, he writes, teaches, and consults on the field of semiconductor microlithography in Austin, Texas.