Journal of
Micro/Nanolithography,
MEMS, and MOEMS

Nanolithography.SPIEDigitalLibrary.org

More systematic errors in the measurement of power spectral density

Chris A. Mack

SPIE.

More systematic errors in the measurement of power spectral density

Chris A. Mack*
Lithoguru.com, 1605 Watchhill Road, Austin, Texas 78703, United States

Abstract. Power spectral density (PSD) analysis is an important part of understanding line-edge and linewidth roughness in lithography. But uncertainty in the measured PSD, both random and systematic, complicates interpretation. It is essential to understand and quantify the sources of the measured PSD's uncertainty and to develop mitigation strategies. Both analytical derivations and simulations of rough features are used to evaluate data window functions for reducing spectral leakage and to understand the impact of data detrending on biases in PSD, autocovariance function (ACF), and height-to-height covariance function measurement. A generalized Welch window was found to be best among the windows tested. Linear detrending for line-edge roughness measurement results in underestimation of the low-frequency PSD and errors in the ACF and height-to-height covariance function. Measuring multiple edges per scanning electron microscope image reduces this detrending bias. [DOI: 10.1117/1.JMM.14.3.033502]

Keywords: power spectral density; autocovariance function; autocorrelation function; height-height covariance function; spectral leakage; line-edge roughness; linewidth roughness.

Paper 15077P received May 12, 2015; accepted for publication Jun. 17, 2015; published online Jul. 14, 2015.

1 Introduction

Line-edge roughness (LER) and linewidth roughness (LWR) in lithography have many detrimental impacts on the devices being made. The nature of these impacts is a function of the nature of the roughness, which includes both the amount of roughness and its frequency content. Low-frequency roughness, occurring over long length scales, behaves like an error in the mean critical dimension or edge position, resulting in feature-to-feature variation.[1] High-frequency roughness gives within-feature variation that we classically recognize as a "rough" feature. The frequency behavior of the roughness is usually characterized by its power spectral density (PSD), which describes how much variance in the feature can be found in each increment of frequency. Knowing the PSD allows one to translate characteristics of the roughness into effects of the roughness. An alternate description is the autocovariance function (ACF), the Fourier transform of the PSD. As many people have shown, measuring the PSD involves important and subtle details.[2–6]

The PSD of a lithographically produced feature can be thought of as existing over a nearly infinite range of frequencies, with zero frequency representing the true average for an infinitely long line or space, and the highest frequencies reaching down to the atomic level. An idealization of reality considers this frequency range as actually infinite, with a continuous PSD describing the roughness. Measurement of the PSD, however, is made by sampling the edge position (in the case of LER) or the linewidth (in the case of LWR) of a finite-length feature. In practice, the sampling distance is considerably larger than the smallest distance where roughness is still definable, and the finite-length of the sampled line is far less than infinitely long (or even of the longest feature being printed). For example, one might measure 512 points separated by a distance of 4 nm, for a total line length of just over 2 μm. The resulting discrete PSD exhibits not only random errors (measuring noise is fundamentally noisy), but systematic biases as well. The finite length of the measured line gives rise to spectral leakage, and the too-large sampling distance results in aliasing.

In my previous work on this topic, a number of important random and systematic errors in PSD measurement were identified and various mitigation strategies were proposed.[2] Here, some additional effects will be investigated. First, the use of different window functions will be explored to understand their impact on spectral leakage. While the use of any reasonable window function provides much better results than the uniform (rectangular) window commonly employed, many different windows have been proposed for many different PSD applications. A range of window functions will be evaluated here for the LER/LWR application. Next, the effect of detrending—subtracting the mean or a best-fit line from the measured edge—will be explored. Detrending is often necessary but results in a systematic error in an important part of the roughness spectrum—the low-frequency roughness.

2 PSD Uncertainty

Measuring roughness is noisy business. Since roughness is by definition random, every measured feature will be different, producing a different PSD. It is well-known that the measured PSD of random noise (sometimes called the Wiener spectrum) has a relative standard deviation of one at each frequency, that is

$$\frac{\sqrt{\mathrm{var}(\mathrm{PSD}(f))}}{\langle \mathrm{PSD}(f) \rangle} = 1. \tag{1}$$

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

1932-5150/2015/$25.00

Further, the variance of the PSD at a given frequency is independent of the PSD at all other frequencies.[7] Since the PSD at each frequency is the sum of the squares of the real and imaginary parts of the Fourier transform of the roughness, a normal distribution of roughness will produce a sampling distribution for the PSD that is χ^2 with two degrees of freedom (that is, an exponential distribution with parameter 1/2). At the extremes of zero and the Nyquist frequency, there is one degree of freedom since the imaginary part of the Fourier transform will always be zero at these frequencies. If $\mathrm{PSD}_m(f)$ is the measured (estimated) value of the PSD at a specific frequency, and $\langle \mathrm{PSD}_m(f) \rangle$ is the mean value averaged over many realizations of the roughness (that is, the true value), then with ν degrees of freedom,

$$\frac{\nu \mathrm{PSD}_m(f)}{\langle \mathrm{PSD}_m(f) \rangle} \sim \chi^2(\nu). \tag{2}$$

Since the variance of the χ^2 distribution is 2ν, we see that, in general

$$\sqrt{\mathrm{var}[\mathrm{PSD}_m(f)]} = \frac{\langle \mathrm{PSD}_m(f) \rangle}{\sqrt{\nu/2}}. \tag{3}$$

Error bars can easily be established for every measured PSD value. The $(1-\alpha)\%$ confidence interval will be[7]

$$\frac{\nu \mathrm{PSD}_m(f)}{\chi^2_{1-\alpha/2}(\nu)} < \langle \mathrm{PSD}_m(f) \rangle < \frac{\nu \mathrm{PSD}_m(f)}{\chi^2_{\alpha/2}(\nu)}. \tag{4}$$

If 95% confidence interval error bars are desired for a measured PSD value, the critical values for the $\chi^2(\nu = 2)$ distributions are 0.051 and 7.38. The 95% confidence interval for a measured PSD value would then be $(0.271\,\mathrm{PSD}_m(f),\ 39.5\,\mathrm{PSD}_m(f))$, a very wide and very asymmetric range.

Because of this very high uncertainty for any measured PSD value, it is common to measure multiple PSDs and average them. If M PSDs are measured and averaged together, the resulting averaged PSD will have $2M$ degrees of freedom. Thus, the standard deviation of the measurements of the PSD are reduced by $1/\sqrt{(M)}$. As an example, averaging 25 PSDs will result in a 95% confidence interval for each PSD point of $[0.70\,\mathrm{PSD}_m(f),\ 1.55\,\mathrm{PSD}_m(f)]$. If 100 PSDs are averaged, the 95% confidence interval is $[0.83\,\mathrm{PSD}_m(f),\ 1.23\,\mathrm{PSD}_m(f)]$. For large degrees of freedom (>100, for example), the χ^2 distribution approaches a normal distribution with a standard deviation equal to the $mean/\sqrt{(M)}$ An example of a single PSD and an averaged PSD with $M = 25$, each showing error bars, is found in Fig. 1. Because the y-axis is on a logarithmic scale, the lengths of the error bars will be the same for every data point and it is sufficient to plot error bars for one data point. The inclusion of error bars on a measured PSD should be standard practice in order to avoid over interpretation of the data, but unfortunately is quite rare.

Systematic errors in PSD measurement are caused by several factors.[2] 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 (Nyquist) frequency, equal to $1/(2\Delta y)$ for a sampling distance of Δy.

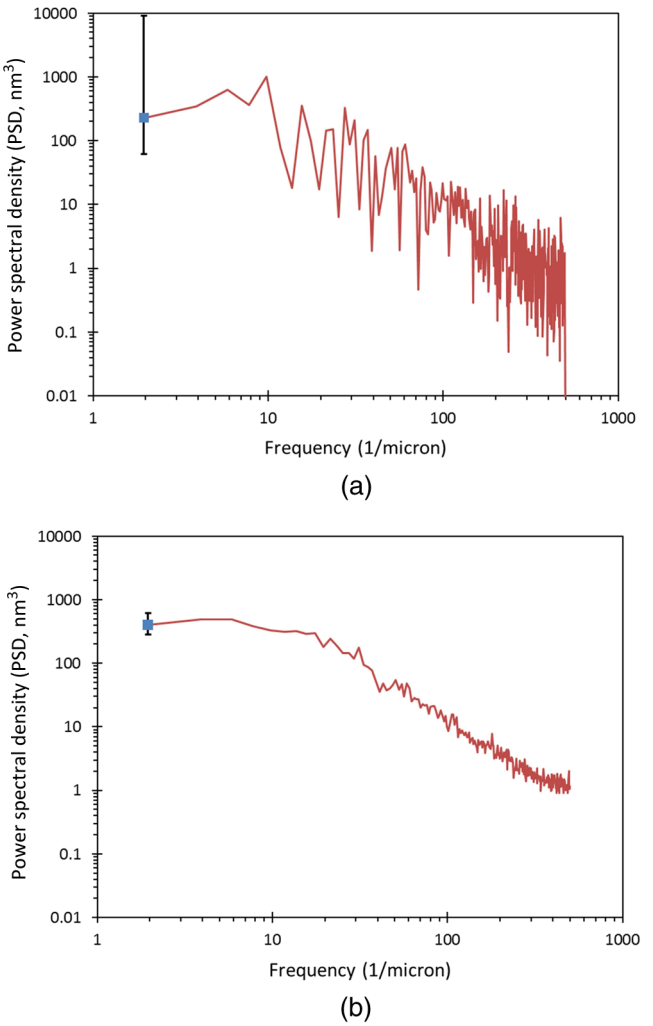

Fig. 1 Examples of (a) a single power spectral density (PSD), and (b) an averaged PSD ($M = 25$), each showing 95% confidence interval error bars displayed for one data point. PSDs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, with leakage and aliasing included in the simulation.

Averaging occurs whenever the measurement spot size is an appreciable fraction of the sampling distance. Scanning electron microscope (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, thus complicating interpretation of a measurement that we hope will give insight into the true roughness behavior of the feature. A thorough understanding of these effects can be used to minimize and/or correct for the systematic errors, resulting in a measured PSD much closer to the actual PSD.[2]

3 Reducing Spectral Leakage through Windowing

Spectral leakage is caused by the finite length of the measurement window, L. Consider the simple but useful case of a stretched exponential ACF,

$$\mathrm{ACF}_c(\tau) = \sigma^2 e^{-(|\tau|/\xi)^{2\alpha}}, \tag{5}$$

where τ is the distance, ξ is the correlation length, α is the roughness exponent, and σ is the true LER/LWR (as opposed to that measured for a finite line length). For $\alpha = 0.5$, the resulting continuous PSD can be analytically derived,[8] leading to the Palasantzas PSD function[9] with a roughness exponent of $H = 0.5$:

$$\mathrm{PSD}_c(f) = \frac{2\sigma^2\xi}{1 + (2\pi f\xi)^2}. \tag{6}$$

As shown previously,[2] the expected value of the sampled PSD (which we will call PSD_d, the discrete PSD) will differ from the true (continuous) PSD as

$$\langle \mathrm{PSD}_d(f)\rangle = \mathrm{PSD}_c(f)(1 + \varepsilon_{\mathrm{alias}})(1 + \varepsilon_{\mathrm{leakage}}), \tag{7}$$

where

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

and $\varepsilon_{\mathrm{alias}} \approx (\pi f\,\Delta y/\sin(\pi f\,\Delta y))^2 - 1$. Thus, the discrete PSD is equal to the continuous PSD modified by two error terms, $\varepsilon_{\mathrm{alias}}$ and $\varepsilon_{\mathrm{leakage}}$. While the above error terms are for the $H = 0.5$ case, they have similar behaviors for other roughness exponents.[2]

Windowing involves multiplying the data set by a window function. If $w(y)$ is the set of measured linewidths (for LWR) or edge positions (for LER), measured at discrete y values that are integer multiples of Δy, then windowing involves weighting the data by a window $g(y)$ before taking the Fourier transform. Standard LER/LWR measurement can be thought of as applying a rectangular measurement window to a long feature: in the region of the line being measured $g(y) = 1$ and outside the region of line being measured $g(y) = 0$.

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.

$$\begin{aligned}\langle \mathrm{PSD}_m(f)\rangle &= \left\langle \left| \int_{-\infty}^{\infty} g(y)(w(y) - \langle w\rangle)\, e^{-i\,2\pi f y}\mathrm{d}y \right|^2 \right\rangle \\ &= G^2(f) \otimes \mathrm{PSD}_c(f),\end{aligned} \tag{8}$$

where $G(f)$ is the Fourier transform of the window function $g(y)$. For the rectangular window of a conventional LER measurement, the continuous PSD is convolved with

$$G^2(f) = \left(\frac{\sin(\pi f L)}{\pi f L}\right)^2. \tag{9}$$

As L becomes large, the spectral window term of Eq. (9) 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. The large sidelobes of the sinc function of Eq. (9) come from the rectangular shape of the data window and result in a large amount of leakage when L is small.

Note that Eq. (9) 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.[2] 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 window convolution term falls off, 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 at least as fast as $1/f^3$.

Here, we consider several popular window functions commonly used in spectral analysis. Table 1 shows the windows studied here, where the window functions must be further normalized by dividing by the root-mean-square value of the window over the line length L. It is important to note that all of the windows are symmetric about $L/2$. Further, all of the windows go to zero at $y = (0, L)$ except the Hamming window.

The simplicity of the Bartlett window allows an analytical approach. The Fourier transform of the Bartlett window gives

Table 1 Common window functions for spectral analysis.

Window	Window function (unnormalized) for $0 < y < L$ (0 otherwise), $\tilde{y} = 2\lvert y - L/2\rvert/L$	c_{window}
Bartlett (Triangle)	$g_{\mathrm{Bartlett}}(y) = 1 - \tilde{y}$	1
Welch	$g_{\mathrm{Welch}}(y) = 1 - \tilde{y}^2$	0.84
Hann (Hanning)	$g_{\mathrm{Hann}}(y) = 1 - \cos(2\pi y/L)$	1.09
Hamming	$g_{\mathrm{Hamming}}(y) = 0.54 - 0.46\cos(2\pi y/L)$	—
Blackman	$g_{\mathrm{Blackman}}(y) = 0.42 - 0.5\cos(2\pi y/L) + 0.08\cos(4\pi y/L)$	1.48
Nuttall	$g_{\mathrm{Nuttall}}(y) = a_0 - a_1\cos(2\pi y/L) + a_2\cos(4\pi y/L) - a_3\cos(6\pi y/L)$ $a_0 = 0.355768,\ a_1 = 0.487396,\ a_2 = 0.144232,\ a_3 = 0.012604$	2.08
Parzen	$g_{\mathrm{Parzen}}(y) = \begin{cases} 1 - 6\tilde{y}^2 + 6\tilde{y}^3 & L/4 < y < 3\ L/4 \\ 2(1-\tilde{y})^3 & y < L/4, y > 3\ L/4 \end{cases}$	1.85

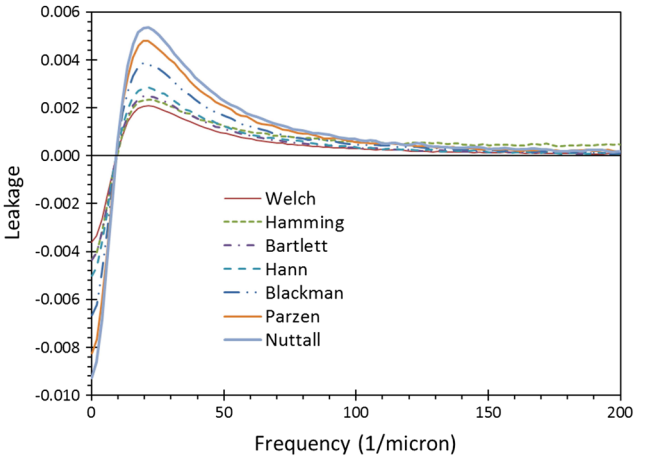

Fig. 2 Calculation of spectral leakage for different window functions. PSDs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, and each curve is generated from the average of at least 10^9 simulated PSDs.

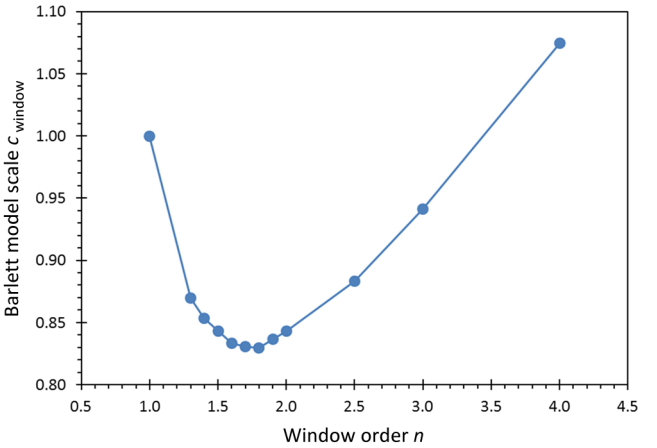

Fig. 3 Varying the window order from the generalized Welch window of Eq. (13) and fitting the resulting leakage to the scaled Bartlett model of Eq. (12) allows one to determine the window that produces the minimum leakage. Best fit c_{window} values were found using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, and the average of 10^{10} simulated PSDs.

$$G^2_{\text{Bartlett}}(f) = \left(\frac{\sin(\pi f L/2)}{\pi f L/2}\right)^4. \tag{10}$$

Since this window term falls off as $1/f^4$, the high frequencies of the PSD will not experience significant leakage. In the continuum limit, the leakage for the Bartlett window can be calculated analytically[2]

$$\varepsilon_{\text{leak,Barlett}} = 12\left(\frac{\xi}{L}\right)^2\left(\frac{3(2\pi f\xi)^2 - 1}{[(2\pi f\xi)^2 + 1]^2}\right). \tag{11}$$

For other windows, simulation can be used to calculate the leakage term. Figure 2 shows the results following the simulation and calculation procedure described in Ref. 2.

First, note that all of these window functions produce much lower leakage than the rectangular window at high frequencies, and lower leakage at low frequencies whenever $L \gtrsim 12\xi$. For the rectangular window, the maximum leakages for $H = 0.5$ are $\pm\xi/L$, which is about ± 0.02 for the case depicted in Fig. 2. Thus, windowing can reduce the leakage by at least an order of magnitude for this case. It is also interesting to see that all of the windows produce leakages that go to zero at high frequency except the Hamming window due to its nonzero endpoints. It is also clear that the best performing window of this set is the Welch window.

To quantify the quality of leakage reduction, we take note that each window (other than Hamming) follows the same basic $\varepsilon_{\text{leakage}}(f)$ behavior. Thus, we can use the analytical result for the Bartlett window and model the leakage as

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

We can now find the value of c_{window} that produces the best fit to the simulated leakage data. Those values are listed in Table 1. For example, the Welch window results in 84% of the leakage of the Bartlett window. The Hann and Blackman windows produce about 9% and 48% more leakage, respectively, than the Bartlett window.

Examining the relationship between the Bartlett window and the best window of those investigated here, the Welch window, one can see a generalization of form

$$g_n(y) = 1 - \tilde{y}^n, \quad \tilde{y} = \frac{|y - L/2|}{L/2}. \tag{13}$$

When $n = 1$, we have the triangular Bartlett window. When $n = 2$, we have the parabolic Welch window. Other values of n, then, would produce other windows, which I will refer to as generalized Welch windows. Varying the window order n and determining the best fit c_{window} from Eq. (12) gives results as shown in Fig. 3. A window order of $n = 1.7$–1.8 produces the best leakage suppression, slightly better than the Welch window and 17% better than the Bartlett window. An alternate metric, the RMS leakage over the full frequency range, produces a similar curve with the same best window and a minimum RMS $\varepsilon_{\text{leakage}}$ of 0.0006.

While the results presented in Fig. 3 are for the case of a roughness exponent of $H = 0.5$, simulations were also performed for $H = 0.9$. The same trend as seen in Fig. 3 was found when $H = 0.9$, with the values of c_{window} about 40% higher. A window order of $n = 1.7$–1.8 again produced the best leakage suppression.

4 Impact of Detrending

Using simulation, it is possible to generate random rough features to be analyzed where the exact statistical properties of the feature are known *a priori*.[10] For example, when generating a rough line, the mean linewidth of an infinitely long line is an input to the simulation. When analyzing the simulated rough line, it is possible to take advantage of this knowledge and calculate the PSD of the deviation of the linewidth from the known population mean. But with experimental data this population mean is never known and so must be estimated, generally by using the sample mean. Likewise, a simulated rough edge has a known ideal edge position. But the roughness of an experimental edge must be measured against an assumed edge position, generally taken to be the best-fit line through the data. Subtracting the sample mean

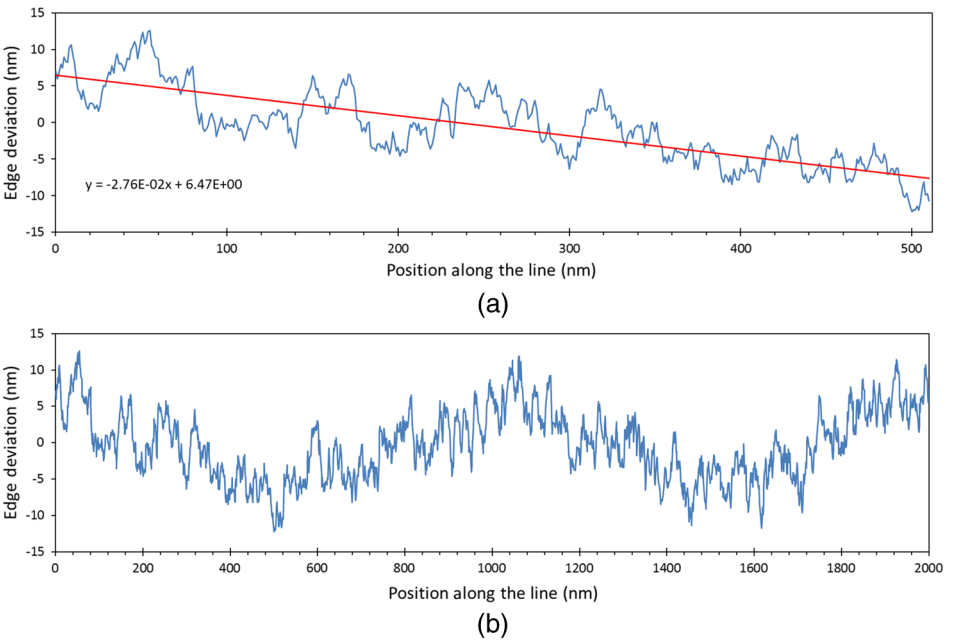

Fig. 4 Example of the impact of detrending on line-edge roughness measurement. (a) The first 25% of a longer line segment, and (b) the full line segment, indicating that low-frequency roughness is removed by linear detrending.

for LWR data and the best-fit linear trend for LER data is called detrending the data. In other applications, such as measuring the roughness of optical components, higher order detrending is sometimes used.[11]

Detrending, while necessary, causes a loss of information. For the PSD, that loss occurs at low frequencies. For example, a segment of a line as shown in Fig. 4(a) is detrended by subtracting out the best-fit line through the data. This is necessary, for example, if the wafer was rotated slightly before measurement in the SEM. The trend, however, could also be part of a low-frequency variation that would be revealed if a longer line segment was measured, as in Fig. 4(b).

The impact of detrending, both sample mean and linear subtraction, can be studied through simulation. Subtracting the sample mean before calculating the PSD has a very simple effect: it forces the zero-frequency PSD to be 0, but otherwise has no impact on the PSD of higher frequencies. Linear detrending has an impact that is felt over the course of all the low-frequency PSD values. Figure 5 shows an example of a PSD with and without linear detrending (a similar behavior was found for larger roughness exponents as well). The difference, that is the bias caused by detrending, can be explained by a simple empirical model based on the major trend observed in Fig. 5(b)

$$\langle \mathrm{PSD}_{\mathrm{linear}}(f)\rangle = \langle \mathrm{PSD}_d(f)\rangle(1-\varepsilon_{\mathrm{linear}}),$$
$$\varepsilon_{\mathrm{linear}} = \frac{1}{\pi(fL)^2}. \tag{14}$$

As can be seen, the biggest impact of linear detrending is on the lowest frequencies. The lowest measureable frequency of the PSD occurs at $f = 1/L$, so that the linear detrending error at this frequency is $1/\pi$, or about 30%. This will be a noticeable bias if 10 or more PSDs are averaged together. The second lowest frequency will have a linear detrending bias of $1/4\pi$, or about 8%. This bias would not be noticed unless 100 or more PSDs are averaged together.

The above results assumed a rectangular data window for PSD calculation. When a higher-order window such as the Welch window is used, the impact of linear detrending is less significant at the higher frequencies (Fig. 6). Simulations suggest the following empirical model using the Welch or other higher-order windows:

$$\varepsilon_{\mathrm{linear}} = \frac{1}{\pi(fL)^5}. \tag{15}$$

Thus, when using a reasonable data window, only the lowest frequency ($f = 1/L$) is appreciably affected by linear detrending. However, at this lowest frequency the use of the Welch window or other data windows has no effect on the detrending bias.

The impact of detrending can be further reduced if multiple features are detrended together. For example, it is common to capture an SEM image with multiple features. If LER is to be extracted from each of the feature edges, a simple approach would be to fit individual straight lines to each edge, then measure the LER after individual detrending. This produces the $1/\pi$ bias for PSD($f = 1/L$). Alternately, all of the edges can be detrended together, assuming that every true edge is parallel (a very common case). Thus, one slope is found that best fits every edge in the image. This is equivalent to taking out one rotational error for the entire SEM image. As a result, the linear detrending bias is reduced by the number of edges in the image being detrended together (true whether a rectangular or other data window is used). Thus, for a rectangular window we have

Thus, for a rectangular window we have for k edges per image,

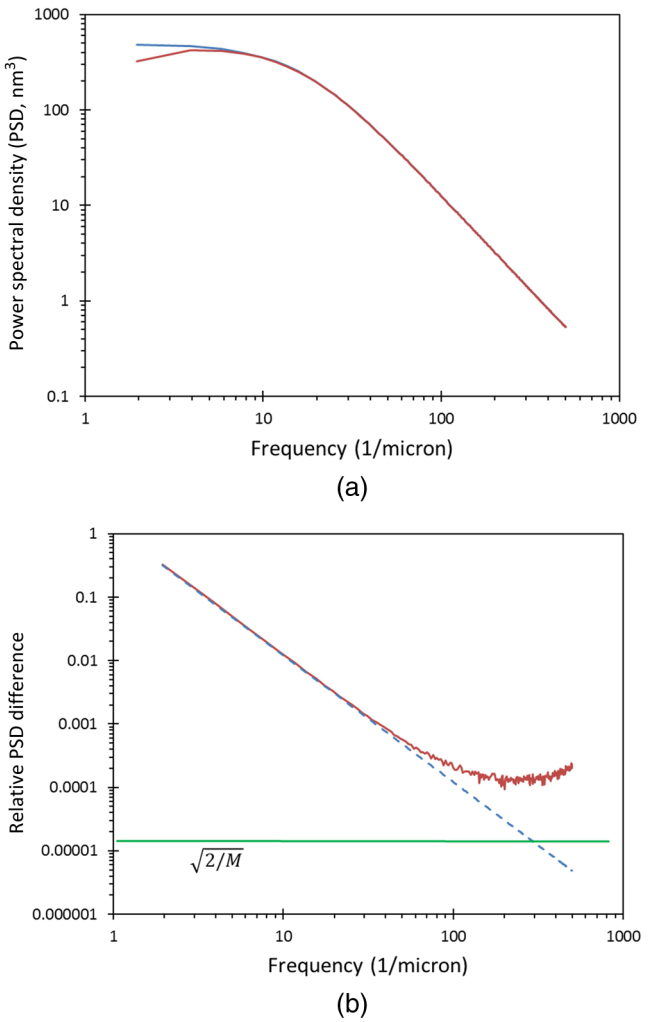

Fig. 5 Impact of linear detrending on the resulting PSD using a rectangular data window: (a) PSD without detrending (upper, blue curve) compared to the PSD with linear detrending (lower, red curve), and (b) the relative difference between the original and detrended PSDs (that is, $\varepsilon_{\text{linear}}$), with the noise floor indicated by the horizontal line and the model of Eq. (14) depicted as the dashed line. PSDs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, and the Palasantzas PSD function with $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, $M = 10^{10}$ PSDs averaged together, and aliasing was turned off.

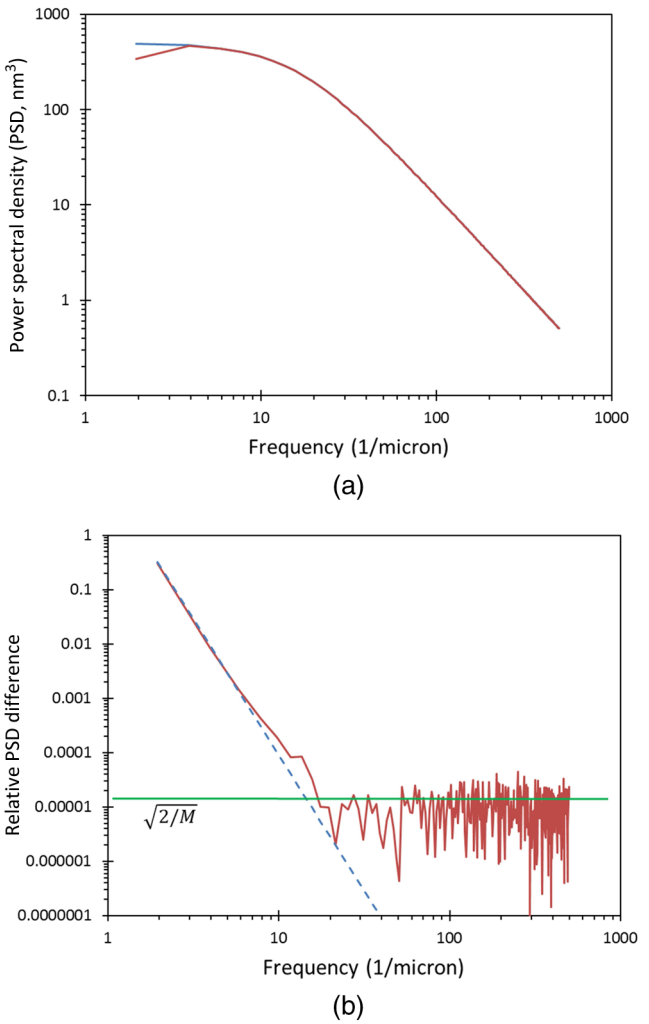

Fig. 6 Impact of linear detrending on the resulting PSD using the Welch data window: (a) PSD without detrending (upper, blue curve) compared to the PSD with linear detrending (lower, red curve), and (b) the relative difference between the original and detrended PSDs, with the noise floor indicated by the horizontal line and the model of Eq. (15) depicted as the dashed line. PSDs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, and the Palasantzas PSD function with $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, $M = 10^{10}$ PSDs averaged together, and no aliasing. Similar results were observed for higher roughness exponents.

$$\varepsilon_{\text{linear}} = \frac{1}{k\pi(fL)^2}. \tag{16}$$

If four or more edges are linearly detrended per image, the impact of this detrending bias will be small enough to be neglected in most circumstances.

5 Impact of Detrending on the ACF and Height-Height Covariance Function

While it is common to use the PSD to describe and evaluate the frequency dependence of measured roughness, the ACF is sometimes used as well. The ACF can be calculated from the Fourier transform of the PSD, but it can also be calculated directly from sampled data. As we shall see, detrending has a significant impact on the directly measured ACF.

The ACF of the feature edge position (or feature width) is defined as

$$\text{ACF}(\tau) = \langle (w(s) - \langle w \rangle)(w(s+\tau) - \langle w \rangle) \rangle, \tag{17}$$

where w is the measured linewidth/edge position, s and $s + \tau$ are the positions where measurements are made along the length of the line, $\langle w \rangle$ is the true mean linewidth/edge position of the feature, and stationarity is assumed. To calculate the ACF directly from sampled data, with lag $\tau = m\Delta y$, position $s = n\Delta y$, and line length $L = N\Delta y$, a common unbiased estimator is

$$\text{ACF}_d(\tau) = \frac{1}{N-m} \sum_{n=0}^{N-1-m} (w(s) - \langle w \rangle)(w(s+\tau) - \langle w \rangle). \tag{18}$$

The fact that the discrete, sampled ACF is an unbiased estimator of the true ACF, i.e., that $\langle \text{ACF}_d(\tau) \rangle = \text{ACF}(\tau)$,

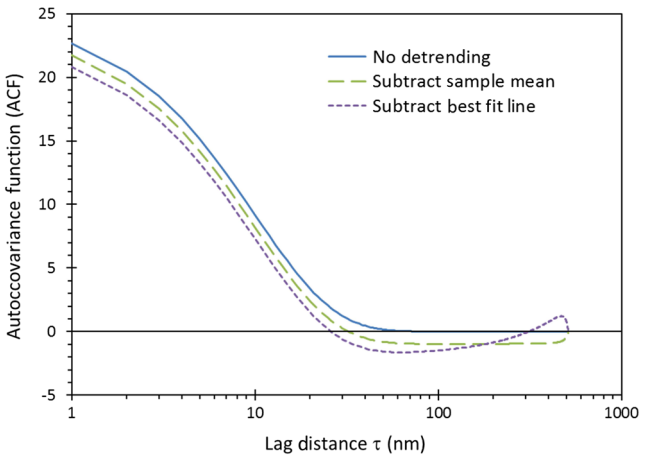

Fig. 7 Impact of mean and linear detrending on the resulting autocovariance function (ACF). ACFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, and the Palasantzas PSD function with $\xi = 10$ nm, $\sigma = 5$ nm, $H = 0.5$, and $M = 10^8$ ACFs averaged together. A logarithmic scale for the lag distance was chosen to make the differences in ACF more noticeable.

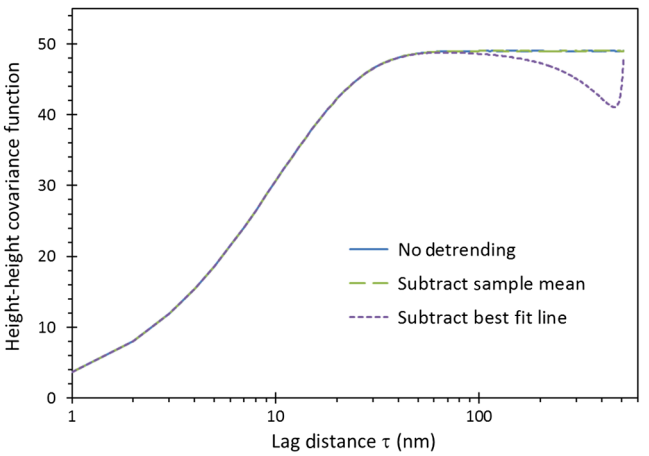

Fig. 8 Impact of linear detrending on the resulting HHCF. HHCFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, $\alpha = 0.5$, and $M = 10^7$ HHCFs averaged together. A logarithmic scale for the lag distance was chosen to make the differences in HHCF more noticeable.

is significant. While the sampled PSD suffers from the significant biases of leakage and aliasing, the discrete ACF does not. Detrending, however, does cause a bias. Detrending involves replacing $\langle w \rangle$ in this equation with either $\bar{w}$ for LWR data or with the best fit line for LER data.

The impact of detrending on the ACF has been studied previously by O'Neill and Walther[12] and Freniere et al.[13] They derived general expressions for the bias caused by mean and linear detrending of the continuous ACF (ACF_c). Applying the case of sample mean subtraction to the exponential ACF [that is, Eq. (5) for $\alpha = 0.5$] and evaluating their expression produces

$$\mathrm{ACF}_{-\mathrm{mean}}(\tau) = \mathrm{ACF}_c(\tau) - \frac{2\sigma^2\xi}{L} \times \left[1 + \frac{\xi}{L} - \left(\frac{\xi}{L-\tau}\right)\left(1 + e^{-\tau/\xi} - e^{-(L-\tau)/\xi}\right)\right]. \tag{19}$$

A graph of the ACF with and without sample mean detrending is shown in Fig. 7. The ACF bias generated by simulation matches essentially exactly with Eq. (19). The bias term is approximately constant until the lag approaches to within a few correlation lengths of L. Thus, for most values of the lag,

$$\mathrm{ACF}_c(\tau) - \mathrm{ACF}_{-\mathrm{mean}}(\tau) \approx \frac{2\sigma^2\xi}{L}\left[1 - \frac{\xi}{L}\right]. \tag{20}$$

Note that the detrended ACF goes negative, even though the true ACF is always positive. For the case of linear detrending, the measured ACF crosses zero twice, giving the false impression that there is a damped periodic variation in the correlation behavior. For a small lag (where the linearly detrended ACF stays positive), the bias for the linearly detrended ACF is about constant and about twice the amount as for the subtraction of the sample mean.

The height-height covariance function (HHCF) is sometimes used in place of the ACF. The HHCF is defined as

$$\mathrm{HHCF}(\tau) = \langle [w(s) - w(s+\tau)]^2 \rangle. \tag{21}$$

For sampled data, the unbiased estimator is

$$\mathrm{HHCF}_d(\tau) = \frac{1}{N-m}\sum_{n=0}^{N-1-m}[w(s) - w(s+\tau)]^2. \tag{22}$$

The HHCF is related to the ACF by

$$\mathrm{HHCF}(\tau) = 2[\sigma^2 - \mathrm{ACF}(\tau)]. \tag{23}$$

Like the ACF, the HHCF does not suffer from the biases of aliasing and leakage that plague the PSD. Further, the HHCF has one important advantage over the ACF: it is not biased by detrending using the mean of the sample data. As Eqs. (21) and (22) make clear, subtracting any value from both $w(s)$ and $w(s+\tau)$ does not impact the calculated HHCF. Looking at Eq. (23), the bias in the ACF from mean detrending is exactly compensated by the bias in σ^2 from mean detrending. The same thing is not true for linear detrending. Like the ACF, the HHCF suffers from a bias, strongest at the long lag distances, when linear detrending is used (Fig. 8). Thus, HHCF is very useful for LWR data since it is unbiased. For LER data, both ACF and HHCF are biased.

Like the PSD, the detrending bias in the ACF and HHCF can be reduced by increasing the number of features that are detrended together. Figure 9 shows the error in the HHCF due to linear detrending. When only one edge is used to find the best fit line, the worst case error is about −32% under these conditions (the magnitude of the linear detrending error is a function of ξ/L). If k edges per image are used to find one best-fit slope, the error is reduced by a factor of k. Thus, an SEM image with four features (and eight edges) can be used to find the HHCF of the LER with only about 4% maximum systematic bias (for the case shown in Figs. 8 and 9).

6 Conclusions

In my previous work, several important strategies were developed for accurately measuring the PSD of rough features:[2]

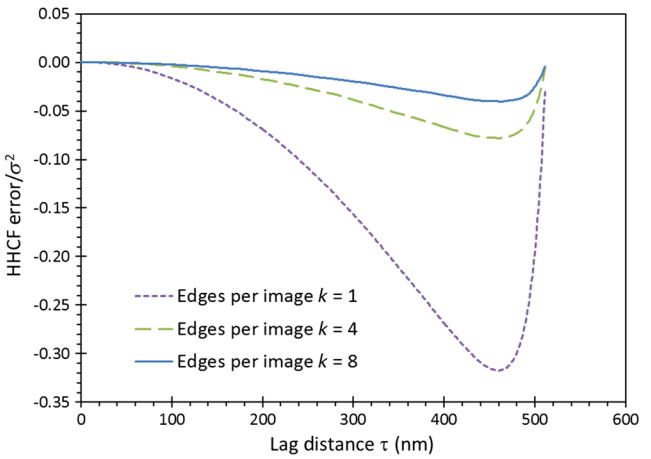

Fig. 9 Impact of linear detrending (with multiple edges per image) on the resulting HHCF. HHCFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, $\alpha = 0.5$, and $M = 10^7$ HHCFs averaged together.

- Average together as many PSDs as possible to reduce random errors (100 averaged PSDs result in 10% random error in the PSD, one standard deviation);
- 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 (the Bartlett, Welch, or similar window) to reduce spectral leakage to small levels;
- Balance aliasing with averaging by optimizing the sampling distance with respect to the interaction spot size of the measurement tool. 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. Reporting only σ is not sufficient.

Here, several further strategies were developed.

- Always put error bars on graphs of measured PSDs to avoid over interpreting the resulting shapes;
- The best data window tested so far is the Welch window (or slightly better, the generalized Welch window with $n \approx 1.7$–1.8)—it should always be used;
- Linear detrending biases the PSD lower at the lowest spatial frequency, corresponding to $f = 1/L$, by about 30%. To effectively eliminate this bias, be sure to detrend four or more edges together in one image;
- The ACF is sensitive to both mean and linear detrending. The HHCF has the advantage of not being sensitive to mean detrending and is thus preferable for LWR analysis. Like the PSD, detrending bias is reduced in direct proportion to the number of features/edges detrended together. Further work is required to more fully understand the implications of the uncertainties and biases of the ACF and HHCF and how to mitigate them;

As the above points make clear, care must be taken in collecting and analyzing roughness data. Since details of the measurement process (line length, sampling distance, number of PSDs averaged together, measurement spot size) and of the analysis process (type of data window and detrending) dramatically affect both the random and systematic errors in the resulting PSD, these details should always be reported when PSD results are presented. Failure to do so makes interpretation of the data almost impossible.

References

1. C. A. Mack, "Analytical expression for impact of linewidth roughness on critical dimension uniformity," *J. Micro/Nanolith. MEMS MOEMS* **13**(2), 020501 (2014).
2. C. A. Mack, "Systematic errors in the measurement of power spectral density," *J. Micro/Nanolith. MEMS MOEMS* **12**(3), 033016 (2013).
3. T. Verduin, P. Kruit, and C. W. Hagen, "Determination of line edge roughness in low-dose top-down scanning electron microscopy images," *J. Micro/Nanolith. MEMS MOEMS* **13**(3), 033009 (2014).
4. 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).
5. L. Azarnouche et al., "Plasma treatments to improve line-width roughness during gate patterning," *J. Micro/Nanolith. MEMS MOEMS* **12**(4), 041304 (2013).
6. A. V. Pret et al., "Resist roughness evaluation and frequency analysis: metrological challenges and potential solutions for extreme ultraviolet lithography," *J. Micro/Nanolith. MEMS MOEMS* **9**(4), 041308 (2010).
7. G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications*, pp. 231–254, Holden-Day, San Francisco (1968).
8. 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).
9. 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).
10. C. A. Mack, "Generating random rough edges, surfaces, and volumes," *Appl. Opt.* **52**(7), 1472–1480 (2013).
11. P. Z. Takacs, "Standardization of methods for extracting statistics from surface profile measurements," *Proc. SPIE* **9173**, 917309 (2014).
12. E. L. O'Neill and A. Walther, "A problem in the determination of correlation functions," *J. Opt. Soc. Am.* **67**(8), 1125–1126 (1977).
13. E. R. Freniere, E. L. O'Neill, and A. Walther, "Problem in the determination of correlation functions. II," *J. Opt. Soc. Am.* **69**(4), 634–635 (1979).

Chris A. Mack developed the lithography simulation software PROLITH and founded the company FINLE Technologies in 1990. He served as a vice president of lithography technology for KLA-Tencor for five years, until 2005. He received the 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.