Systematic Errors in the Measurement of Power Spectral Density

Chris A. Mack
Lithoguru.com, 1605 Watchhill Rd, Austin, TX 78703, E-mail: *chris@lithoguru.com*

Abstract

Measurement of the power spectral density (PSD) of a rough surface or 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 errors. Using both analytical expressions and numerical simulations for the measurement of the PSD of line-edge roughness, three sources of systematic errors are evaluated: aliasing, leakage, and averaging. 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. 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 the extracted roughness exponent to higher values.

Subject Terms: power spectral density, PSD, discrete PSD, aliasing, spectral leakage, line-edge roughness, linewidth roughness, LER, LWR

1. Introduction

Line-edge roughness (LER) and linewidth 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 linewidth (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 systematic biases as well. Thus, it is important to understand the nature and 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 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.[3,4,5,6,7] For other cases, numerically generating synthetic rough edges that are then "measured" and analyzed leads to further insights into errors in PSD measurement.[8,9] In this paper, the properties of the emasured PSD will be examined using these and other analytical tools with the goal of defining and then minimizing the systematic errors present in 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 ($\widetilde{R}$) of the feature edge position (or feature width for the case of two feature edges that are completely uncorrelated) is defined as

Metrology, Inspection, and Process Control for Microlithography XXVII, edited by Alexander Starikov, Jason P. Cain,
Proc. of SPIE Vol. 8681, 86812I · © 2013 SPIE · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2012022

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

$$\widetilde{R}(s,t) = \left\langle \left(w(s) - \overline{w}\right)\left(w(t) - \overline{w}\right) \right\rangle \tag{1}$$

where w is the measured linewidth/edge position, s and t are the positions where measurements are made along the length of the line, $\overline{w}$ is the mean linewidth/edge position of the feature, and $\langle ... \rangle$ 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.

$$\left\langle PSD_d(f) \right\rangle = \frac{\Delta y}{N} \left\langle \left| \sum_{s=0}^{N-1} \left(w(s) - \overline{w}\right) e^{-i2\pi\tau s/N} \right|^2 \right\rangle$$

$$= \frac{\Delta y}{N} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} \widetilde{R}(s,t) e^{i2\pi\tau(s-t)/N} = \Delta y \sum_{m=-(N-1)}^{N-1} \left(1 - \frac{|m|}{N}\right) \widetilde{R}(m) e^{i2\pi\tau m/N} \tag{2}$$

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 right-hand side of equation (2) shows that calculating the PSD from the discrete Fourier transform of linewidth 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 apply to rough features.

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

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),

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

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

$$\frac{\left\langle PSD_d(f) \right\rangle}{2\sigma^2 \xi} = \delta \left(\frac{1}{2} + \mathrm{Re}\left\{ \frac{z}{1-z} \right\} - \frac{1}{N} \mathrm{Re}\left\{ \frac{z - z^{N+1}}{(1-z)^2} \right\} \right) \tag{5}$$

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

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 more in the range of 0.7 – 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, equation (5) is awkward to deal with and is evaluated numerically. Some simplifications, however, will make the analytical PSD_d more convenient. Taking each term separately,

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

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

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

Thus, the discrete PSD becomes

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

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

$$\left\langle PSD_d(f)\right\rangle\approx\frac{2\sigma^2\xi}{1+(2\pi f\xi)^2}\left(1+\varepsilon_{alias}\right)\left(1+\varepsilon_{leakage}\right) \tag{9}$$

where $\varepsilon_{leakage}=\left(\frac{\xi}{L}\right)\left(\frac{(2\pi f\xi)^2-1}{(2\pi f\xi)^2+1}\right)$ and $\varepsilon_{alias}=\left(\frac{\pi f\Delta y}{\sin(\pi f\Delta y)}\right)^2-1$

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

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

shown below, and is captured by the term ε_{alias}. The aliasing term ε_{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 $\varepsilon_{leakage}$ varies from a low of $-\xi/L$ at $f = 0$ to a high approaching ξ/L at the highest frequencies, passing though zero at a frequency corresponding to the correlation length. The term "leakage" refers to the impact of measuring within a finite window (that is, a finite length of the line being measured), resulting in a localized spreading of frequency components. As equation (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_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 to the correlation length.

Thus, there are two sources of difference between the continuous and discrete PSDs: the first is due to the nonzero value of ξ/L (as captured in the $\varepsilon_{leakage}$ term), and the second is due to the nonzero value of Δy (as captured in the ε_{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 equation (3) with $\alpha = 0.5$, so that the true PSD is the continuous PSD of equation (4). Under this assumption, differences between the discrete and continuous PSD are the result of sampling (the non-zero 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 this $\alpha = 0.5$ case) at the Nyquist frequency. Figure 1 shows plots comparing the discrete PSD [equation (5)], the continuous PSD [equation (4)], and the approximation to the discrete PSD [equation (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 as well, as shown in Figure 2.

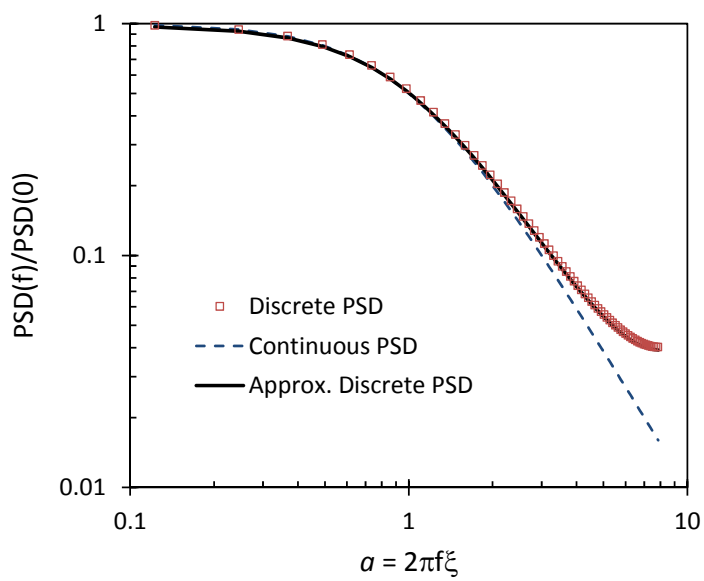
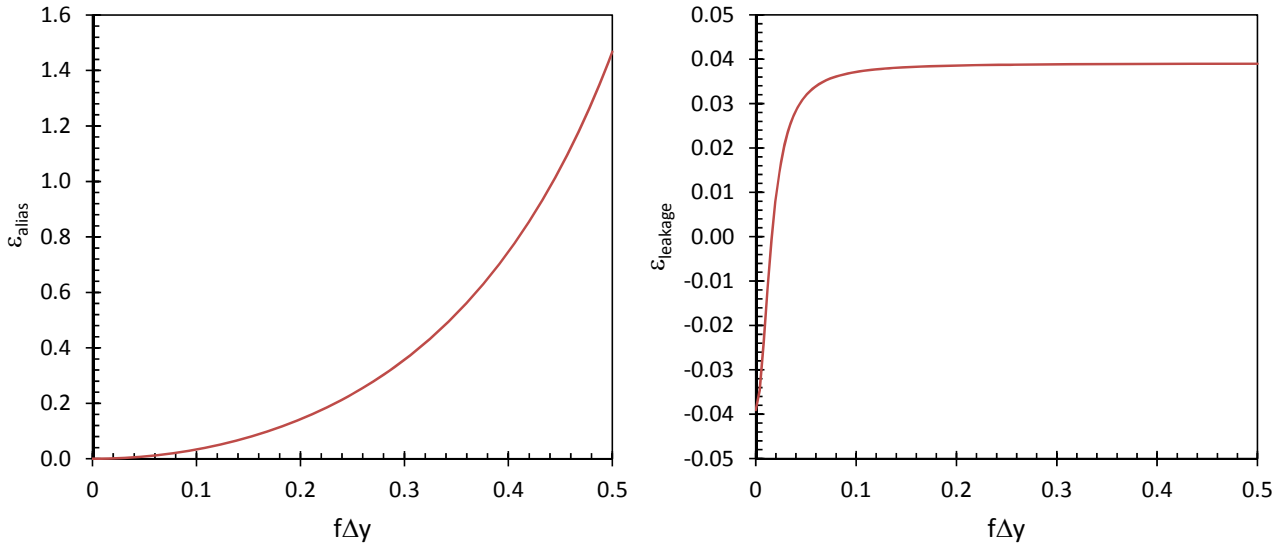


Figure 1. Plots of the discrete PSD (symbols), the continuous PSD (dashed line), and the approximation to the discrete PSD found in equation (9) (solid line), using $N = 128$, $\xi = 6.4$ nm, and $\Delta y = 2.56$ nm.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms



(a) (b)

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

Equation (9) was derived for the single case of $\alpha = 0.5$. However, the form of equation (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 power spectral density (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 $f_s = 1/\Delta y$ = the sampling frequency, and k is any 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]

$$\left\langle PSD_d(f)\right\rangle = PSD(f) + \sum_{k=1}^{\infty}\left[PSD(kf_s - f) + PSD(kf_s + f)\right] \tag{10}$$

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

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

The infinite summation converges to an analytical result:

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

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

While it is valuable to confirm using Kirchner's method the interpretation of ε_{alias} in equation (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 other PSD functions. Consider the Palasantzas extension of the PSD function of equation (4) to other roughness exponents:[13]

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

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 $d = 1$,

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

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

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

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

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

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

The results of the numerical calculations of $\varepsilon_{alias}(f)$ for $0.5 \le H \le 1$ can be fit extremely well to a simple empirical expression. Letting $\varepsilon_{0.5}$ be the analytical aliasing term for the case of $\alpha = 0.5$ [that is, equation (12)],

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

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

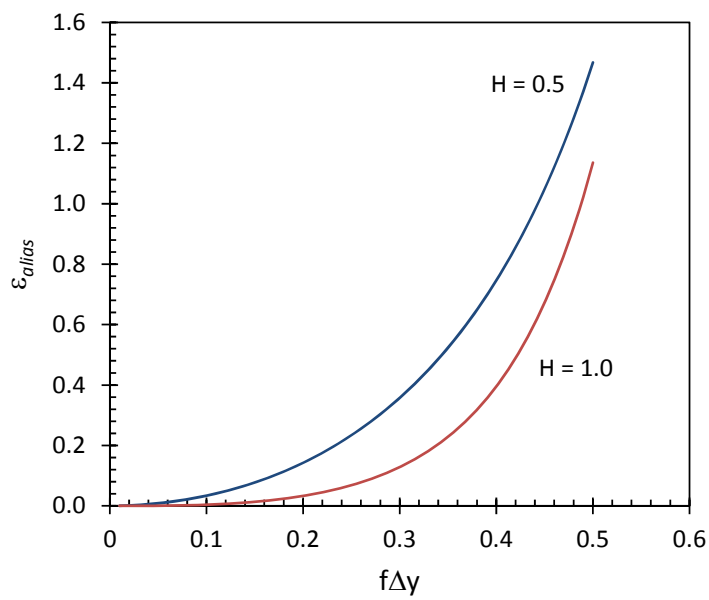


Figure 3. Calculations of ε_{alias} using equation (16) for two different roughness exponents. Results for $0.5 < H < 1.0$ producing aliasing between these two curves.

5. Using Simulation to Determine Leakage and Aliasing

The Hiraiwa and Nishida discrete PSD function of equation (5), as simplified in equation (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 equation (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 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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

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 $L < L_s$. Letting Δs be the simulation grid size, aliasing will occur when the metrology sampling distance $\Delta y > \Delta s$. 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 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 Figure 4 shows, 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 continuous PSDs) when $L = L_s$ and $\Delta y = \Delta s$.

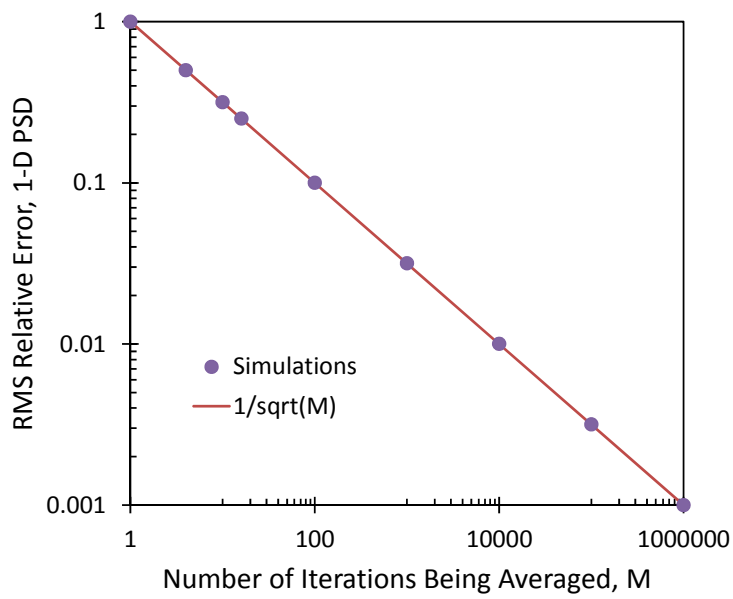


Figure 4. Convergence of the numerically generated PSD to the input PSD as a function of the number of trials being averaged together (σ = 5 nm, ξ = 10 nm, α = 0.5, Δy = 1 nm, N = 1,024). The standard $1/\sqrt{M}$ convergence trend is shows as the solid line, with simulations shown as the symbols. The 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.)

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

$$\left\langle PSD_{simulation}(f)\right\rangle = PSD(f)\left(1+\varepsilon_{leakage}\left(1+\varepsilon_{alias}\right)\right) \tag{18}$$

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) to the derived expression for $\varepsilon_{leakage}$ in equation (9) for the case of α = 0.5. As can be

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

seen, the simulations and the derived analytical expression match extremely well until the very highest frequencies. At the Nyquist frequency, the simulated $\varepsilon_{leakage}$ is higher by about 9%, resulting in a difference in the simulated and predicted PSD of less than 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.

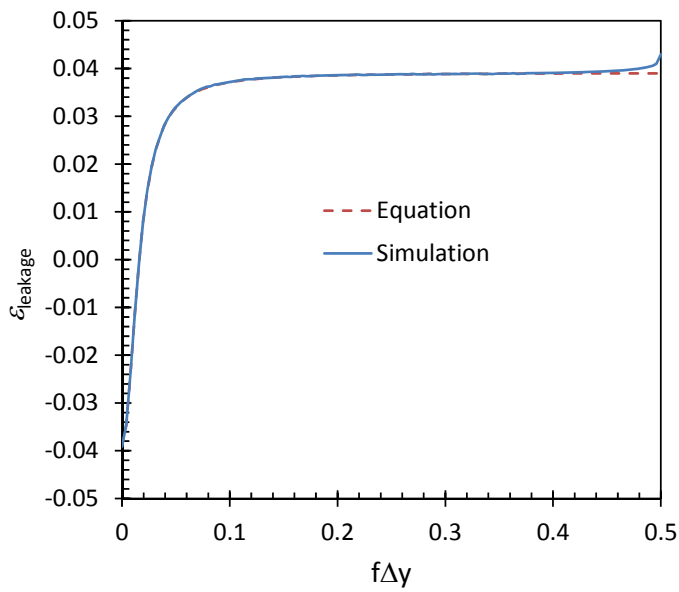


Figure 5. Plots of $\varepsilon_{leakage}$, from equation (9) and from simulations with aliasing turned off, using $\alpha = 0.5$, $N = 256$, $\xi =$ 10 nm, and $\Delta y = 1$ nm.

Likewise, leakage can be turned off by setting $L_s = L$ and aliasing can be turned on by setting $\Delta s = \Delta y/N_{sim}$. Comparison to the analytical aliasing term, however, is complicated by the fact that the analytical result assumes a continuum, that is $N_{sim} \rightarrow \infty$. 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 ε_{alias} calculated from the measured discrete PSD. The results are shown in Figure 6 (10 million simulations per curve). As N_{sim} increases, the aliasing converges to the continuum result from equation (9). The rate of convergence is well described by

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

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 $1/N_{sim}$. Based on this result, the simulations below will use N_{sim} = 128 and the calculated ε_{alias} will be corrected by adding $(2f\Delta y)^2 / N_{sim}$ to give the best approximation to the continuum value of ε_{alias}.

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

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

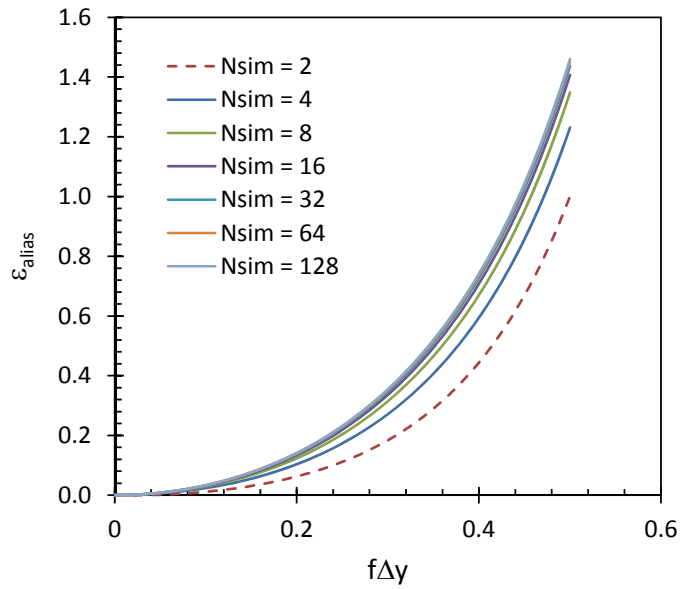


Figure 6. Plots of ε_{alias} from simulations as a function of the ratio of the metrology sampling distance to the simulation grid size (N_{sim}), using $\alpha = 0.5$, $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm.

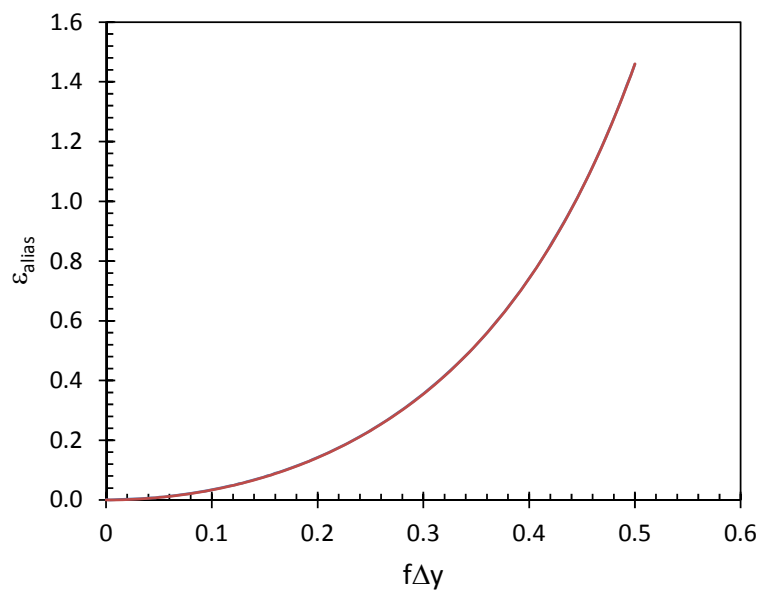


Figure 7. Plots of ε_{alias}, from equation (9) and from simulations with leakage turned off, using $\alpha = 0.5$, $N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm. The two curves are indistinguishable.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

Figure 7 compares ε_{alias} from equation (9) to the results of 8 million simulations. Both curves are plotted on the same graph, but the results are indistinguishable, with differences less than 0.002. The results shown in Figures 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 by equation (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 Figure 8. While the impact of leakage is small for the case of $H = 0.5$, it is much larger for larger roughness exponents.

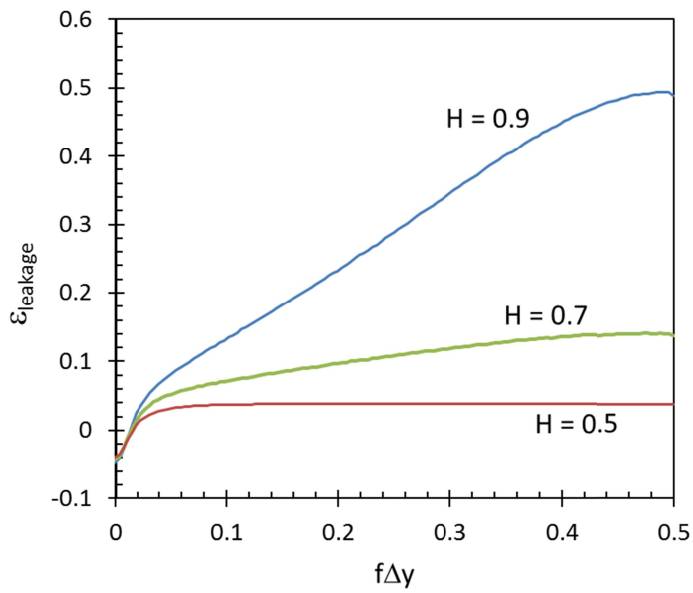


Figure 8. Plots of $\varepsilon_{leakage}$, from simulations, for different values of the roughness exponent H ($N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm).

6. Reducing Leakage with Data Windowing

The above sections describe 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 with data windowing. In this approach, the measurement value $w(s)$ used in equation (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 $g(s) = 1$, outside the region of line being measured $g(s) = 0$. Note that the convolution of this rectangular window with itself produces the $\left(1-|m|/N\right)$ term in equation (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.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

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

where $g(y)$ is assumed 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

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

As L becomes large, window term of equation (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 equation (21) falls off as $1/f^2$ away from the frequency being measured. The PSD, on the other hand, falls off as $1/f^{2H+1}$. 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 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

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

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

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

Since this window term falls off as $1/f^4$, the high frequencies of the PSD will not experience leakage. Other common windows, such as the Welch and Hann windows, have the same behavior.[16] Figure 9 shows simulations of measured PSD using the Bartlett window, extracting the leakage term as before. Note that the resulting leakage is less than 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 Figure 5.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

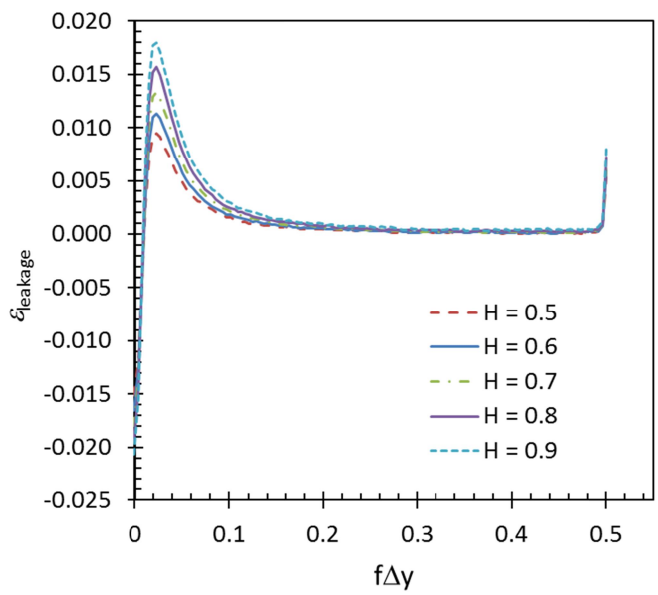


Figure 9. Plots of $\varepsilon_{leakage}$, from simulations using the Bartlett window, for different values of the roughness exponent H ($N = 256$, $\xi = 10$ nm, and $\Delta y = 1$ nm).

7. The Impact of Spatial Averaging

One useful approach to reducing the effects of aliasing is through averaging. If the spacing between measurements is Δy, the measurement can be (and usually is) the average linewidth or edge position over some range η. If $\eta = 0$ then 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 rectangle shape (straight averaging over the distance η), the Fourier transform is a sinc function, giving

$$PSD_{d-avg}(f) = PSD_d(f)\left(\frac{\sin(\pi f \eta)}{\pi f \eta}\right)^2 \tag{24}$$

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) width η. The impact of this averaging can be seen in Figure 10 using simulation, and is 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.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

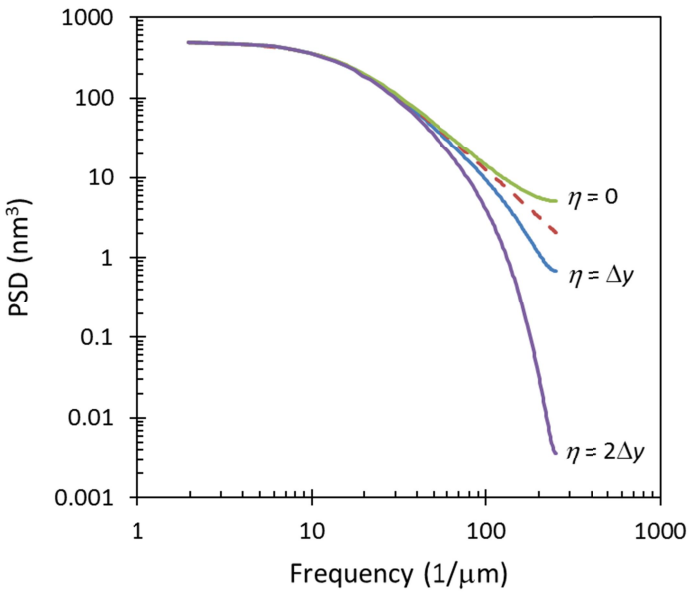


Figure 10. Simulations of the impact of averaging on the measured PSD. The FWHM of the Gaussian measurement signal (η) is varied from 0 to twice the sampling distance. The continuous PSD (without aliasing, leakage or averaging) is shown as the dotted line ($N = 256$, $\xi = 10$ nm, $H = 0.5$, $\Delta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

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 – 6 nm wide, depending on the electron energy. If the sampling distance is set to 4 nm (a commonly recommended value), then averaging would occur over a distance of 1 – 1.5Δy. As Figure 10 shows, the result will be a PSD with suppressed high-frequency power and that appears to have a higher value of the roughness exponent.

It will be convenient to account for this averaging effect by adding an averaging error term ε_{avg}.

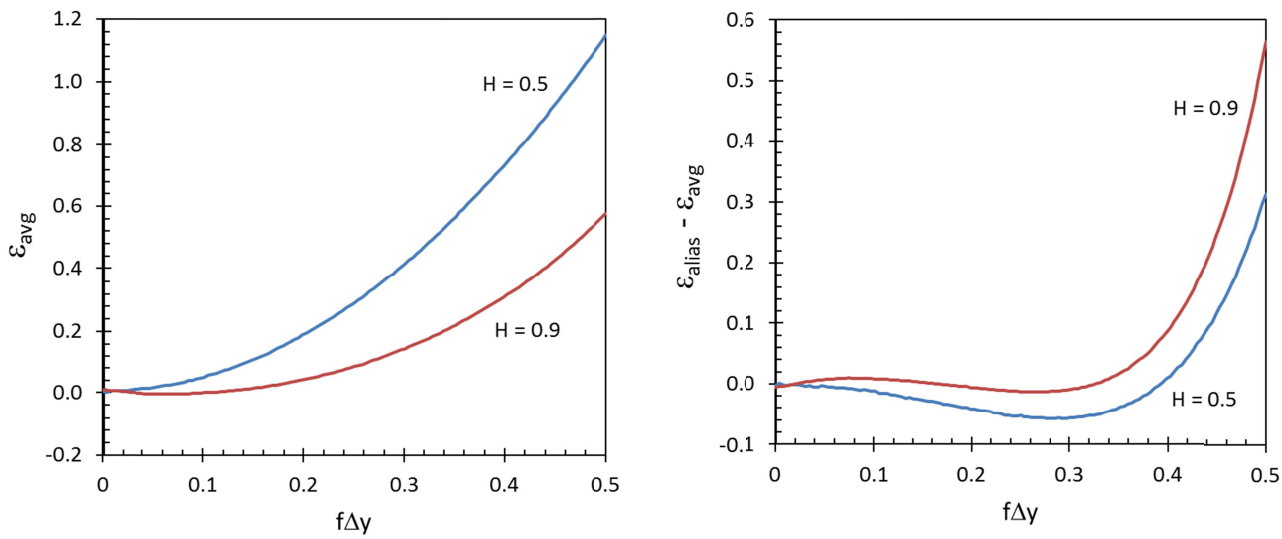
$$\langle PSD_d(f)\rangle = PSD(f)\left(1+\varepsilon_{alias}-\varepsilon_{avg}\right)\left(1+\varepsilon_{leakage}\right) \tag{25}$$

Simulations can be used to determine the magnitude and frequency dependence of ε_{avg}. For example, when $H = 0.5$, Gaussian averaging is well approximated by

$$\varepsilon_{avg} \approx \frac{0.927\eta}{\Delta y}\left(\pi f \Delta y\right)^2 \tag{26}$$

Figure 11 shows simulation results when $\eta = \Delta y/2$. A sampling distance equal to about twice the Gaussian FWHM provides a nearly optimum amount of averaging to reduce the aliasing. As the graphs show, this setting of sampling distance allows the averaging to counteract the aliasing quite well for the lower 75% of the frequency range. Note also that the Kirchner method can be accommodated to calculate $\varepsilon_{alias} - \varepsilon_{avg}$.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms



(a) (b)

Figure 11. Plots of (a) ε_{avg}, and (b) $\varepsilon_{alias} - \varepsilon_{avg}$, using $N = 256$, $\xi = 10$ nm, $\Delta y = 2$ nm, and $\eta = 1$ nm for a Gaussian-shaped measurement signal.

8. 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. 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. A thorough understanding of these effects can be used to minimize and/or correct for the systematic errors.

Three 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 paper 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. Finally, simulation has been used to generate and measure random rough edges and 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 error in 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)
- Use data windowing (using the Bartlett, Welch, or similar window) to reduce spectral leakage to negligible levels
- The non-zero spot size of the measurement signal produces averaging that counters aliasing, and produces an optimum balance when the sampling distance is set to about twice the spot size FWHM

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms

- Extraction of three PSD parameters (σ, ξ, and H) is essential for understand line-edge roughness. The systematic biases in PSD measurement make accurate measurement of the roughness exponent especially difficult

Correcting the measured PSD for systematic biases is possible using the techniques developed above, but requires that η, the measurement signal FWHM, be known. Alternatively, if the sampling distance is set to twice η, the PSD parameters can be extracted in a straightforward manner by using the lower 75% of the frequency range. In either case, measuring the PSD with a SEM that has an unknown value of η produces a PSD with unknown biases.

An important error source not discussed in this work is image noise, resulting in measurement noise in the edge position/linewidth $w(s)$. This error source is random, but causes a systematic error in the PSD.[1,17] Further work 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, R. Steffen, H. Kawada and T. Iizumi, "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", *Journal of Applied Physics*, **106**, 074905 (2009).
[4] A. Hiraiwa and A. Nishida, "Spectral analysis of line edge and line-width roughness with long-range correlation", *Journal of Applied Physics*, **108**, 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 (Oct-Dec, 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 (Apr-Jun, 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 (Jul-Sep, 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, G. P. Patsis, L.H.A. Leunissen, and E. Gogolides, "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 (Oct-Dec, 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, (Apr–Jun, 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. Accoust. Soc. Am*. **83**(1), 78-92 (Jan, 1988).
[15] C. A. Mack, "Generating Random Rough Edges, Surfaces, and Volumes", Applied Optics, **52**(7), 1487–1496 (2013).
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in Fortran 77: The Art of Scientific Computing, second edition, Cambridge University Press (Cambridge: 1992), p. 545 - 549.
[17] A. Hiraiwa and A. Nishida, "Statistical- and image-noise effects on experimental spectrum of line-edge and line-width roughness", *Journal of Micro/Nanolith. MEMS and MOEMS*, **9**(4), 041210 (Oct–Dec, 2010).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 05/06/2014 Terms of Use: http://spiedl.org/terms