Biases and uncertainties in the use of autocovariance and height–height covariance functions to characterize roughness

Chris A. Mack

Biases and uncertainties in the use of autocovariance and height–height covariance functions to characterize roughness

Chris A. Mack[a)]

Lithoguru.com, 1605 Watchhill Rd., Austin, Texas 78703

Measuring the frequency response of roughness is necessary in many applications, leading to the common use of the power spectral density (PSD) of the roughness. But biases and random uncertainties in the PSD have led some to explore the use of the autocovariance function (ACF) and the height–height covariance function (HHCF) instead. These functions also entail systematic biases and random uncertainties when applied to measured roughness, requiring detailed characterization. A combination of analytical derivations and simulations of rough edges have led to a thorough characterization of these biases and uncertainties for the measurement of line-edge and linewidth roughness of lithographically produced features. The results show that ACF estimation is problematic, but that HHCF estimation is a reasonable alternative to PSD analysis under conditions typical of linewidth roughness measurement in the semiconductor industry. © 2016 American Vacuum Society. [http://dx.doi.org/10.1116/1.4961445]

I. INTRODUCTION

Line-edge roughness (LER) and linewidth roughness (LWR) in lithography are increasingly important to semiconductor manufacturing. As feature sizes shrink, the impact of feature roughness on device performance grows. The nature of these device impacts is a function of the nature of the roughness, which includes both the magnitude of roughness and its frequency content. Low frequency roughness, occurring over relatively long length scales, behaves like an error in the mean critical dimension (CD) or edge position, resulting in feature-to-feature variation often called local CD uniformity or local edge placement error.[1] High frequency roughness gives within-feature variation that impacts the electrical behavior of a feature in a very different way.

The frequency behavior of roughness can be characterized in several ways: by its power spectral density (PSD), by its autocovariance function (ACF), and by its height–height covariance function (HHCF). All of these functions are directly related to each other, so that knowledge of one allows determination of the other two. But perfect knowledge of any of these functions is impossible, and the practitioner must estimate these functions from available measured data. Estimation of a roughness function (PSD, ACF, or HHCF) from a finite-length feature sampled with a finite number of measurements produces both random and systematic errors in the estimated function. The magnitude of these errors and their impact on the interpretation of the function depends on which function is being estimated. Thus, an obvious question arises: when characterizing roughness using measurements, is one function preferred over the other two?

Estimating the PSD directly from roughness data (called nonparametric estimation) produces systematic errors caused by leakage and aliasing, and random errors with the PSD's standard deviation equal to its mean value.[2–7] While leakage can be sufficiently minimized through the use of data windowing (also called data tapering), aliasing is only partially mitigated through the high-frequency averaging inherent in any measurement of roughness.[2,3] Detrending (subtracting off the sample mean for LWR and subtracting off a best-fit line for LER) creates further systematic errors in the low-frequency PSD.[3]

The power spectral density is the Fourier transform of the autocovariance function. Thus, the ACF could be estimated from the estimated PSD. However, the ACF can be better estimated directly from the roughness data, producing an estimate with different biases and uncertainties. As will be described below, ACF estimation is even more problematic than PSD estimation. The HHCF is another important (though underutilized) tool for characterizing roughness and a viable alternative to the use of the ACF or the PSD.[8–11] In this paper, both simulation and analytically derived expressions will be used to characterize the systematic and random errors found in the estimation of the ACF and HHCF for LER and LWR. While the examples presented will be taken from the field of semiconductor lithography, the results have wide application to many autoregressive processes, including time series analysis.

II. ACF UNCERTAINTY

Measuring roughness is noisy business. The power spectral density of a single rough feature has a standard error at each frequency equal to the PSD at that frequency (i.e., the relative uncertainty is 100%). Many PSDs measured from features generated by the same process are often averaged to lower this uncertainty. Likewise, the autocovariance function and the height–height covariance function can have large variances for a single feature.

Consider first the ACF of the feature edge position (or feature width), defined as

$$\text{ACF}(s, \tau) = \langle (w(s) - \langle w \rangle)(w(s + \tau) - \langle w \rangle) \rangle, \qquad (1)$$

where w is the measured linewidth/edge position, s and $s + \tau$ are the positions where measurements are made along the

---

a)Electronic mail: chris@lithoguru.com

length of the line, $\langle w \rangle$ is the true mean linewidth/edge position of the feature, and $\langle ... \rangle$ means an average over many instances of rough features. Throughout this paper, all features will be assumed to have the properties of stationarity (that is, the statistical properties of the feature, such as mean, variance, covariance, etc., remain constant along the length of the feature) and ergodicity (the statistical properties of a feature, such as mean, variance, covariance, etc., remain constant from feature to feature). Under these assumptions, the ACF is only a function of τ (not s), and averaging over many features can be replaced by averaging over the length of one feature. [Although terminology is not uniform, it is common to call ACF(τ)/ACF(0) the *autocorrelation function*. To avoid confusion, this paper will only refer to the autocovariance function as defined by Eq. (1).]

To estimate the ACF directly from sampled (discrete) data, with sampling distance Δy, lag $\tau = m\Delta y$, position $s = n\Delta y$, line length $L = N\Delta y$, and a total of N measurements along the line, a common estimator is

$$R_d(\tau) = \frac{1}{K} \sum_{n=0}^{N-1-m} \big(w(s) - \langle w \rangle \big) \big(w(s+\tau) - \langle w \rangle \big), \qquad (2)$$

where $K = N - m$ produces an unbiased estimator, and $K = N$ produces the more commonly used biased estimator. The reason why a biased estimator is more popular will be explained below. For the case of continuous data, the ACF estimator becomes

$$R_c(\tau) = \frac{1}{T} \int_0^{L-\tau} \big(w(s) - \langle w \rangle \big) \big(w(s+\tau) - \langle w \rangle \big)\, ds, \qquad (3)$$

where the unbiased and biased estimators use $T = L - \tau$ and $T = L$, respectively.

A. Variance of ACF estimators

For the continuous data case, the variance of the ACF estimate can be calculated from[12,13]

$$\mathrm{var}(R_c(\tau)) = \frac{1}{T^2} \int_{-(L-\tau)}^{L-\tau} (L - \tau - |r|)\big(\mathrm{ACF}^2(r) + \mathrm{ACF}(r+\tau)\mathrm{ACF}(r-\tau)\big)\, dr, \qquad (4)$$

where the distribution of line edge positions or linewidths is assumed to be Gaussian and the mean is assumed to be known. Consider an extremely common model for the autocovariance function of roughness, an exponential

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

where σ^2 is the true variance of the line edge/linewidth and ξ is the correlation length. For the discrete case, this model is also known as the autoregressive model of order 1, AR(1). Using this model in Eq. (4), the variance of the ACF estimator is[13,14]

$$
\begin{aligned}
\frac{\mathrm{var}(R_c(\tau))}{\sigma^4} &= \frac{\xi^2}{2T^2} \Big\{ 2e^{-\alpha(1-y)} + \alpha(1-y) - 1 + e^{-\alpha y}\big[\alpha(1-2y) - 1 + \alpha^2 y(1 - 3y/2)\big] \Big\} \quad \text{for } 0 \le y \le 1/2 \\
&= \frac{\xi^2}{2T^2} \Big\{ e^{-\alpha(1-y)} + \alpha(1-y) - 1 + e^{-\alpha y}\big[\alpha^2(1-y)^2/2\big] \Big\} \quad \text{for } 1/2 \le y \le 1,
\end{aligned}
\qquad (6)
$$

where $\alpha = 2L/\xi$ and $y = \tau/L$. Note that in the classic text by Jenkins and Watts,[13] their Eq. (5.3.23) is in error for $\tau \ge L/2$ ($y \ge 1/2$). The correct equation is shown here, and is correct in the original reference by Fuller.[14] Note also that

$$\langle R_c(\tau) \rangle^2 = \frac{\sigma^4 \xi^2}{2T^2} \Big\{ e^{-\alpha y}\big[\alpha^2(1-y)^2/2\big] \Big\} = \left[\frac{L-\tau}{T} \mathrm{ACF}(\tau) \right]^2. \qquad (7)$$

The discrete case can similarly be evaluated by changing Eq. (4) from an integral to a summation. Defining $\rho = e^{-\Delta y/\xi}$, the autoregressive coefficient for lag 1, this results in summations of powers of ρ, which can be evaluated analytically. The result is

$$
\begin{aligned}
\frac{\mathrm{var}(R_d(m))}{\sigma^4} &= \frac{1}{K^2} \left(\frac{2\rho^2}{(1-\rho^2)^2} \right) \Big\{ 2\rho^{2(N-m)} + (N-m)\left(\frac{1-\rho^4}{2\rho^2} \right) - 1 + \rho^{2(m-1)}\big[(1-\rho^2)(N - 2m + 1) - 1\big] \Big\} \\
&\quad + \frac{1}{K^2} \rho^{2m}\big[(N-m)(2m-1) - m(m-1)\big] \quad \text{for } 0 \le m \le N/2 \\
&= \frac{1}{K^2} \left(\frac{2\rho^2}{(1-\rho^2)^2} \right) \Big\{ \rho^{2(N-m)} + (N-m)\left(\frac{1-\rho^4}{2\rho^2} \right) - 1 \Big\} + \left(\frac{N-m}{K} \rho^m \right)^2 \quad \text{for } N/2 \le m \le N.
\end{aligned}
\qquad (8)
$$

In the limit of small Δy, Eq. (8) converges to Eq. (6).

These equations were also checked against simulations. A method for generating randomly rough features with a given autocorrelation behavior has been previously described.[15] This method was used to simulate rough features, creating an instantiation of a zero-mean random variable $w(s)$. Using Eq. (2), both biased and unbiased estimators for the discrete ACF were calculated from a simulated rough feature. Repeating this process on the order of 10^8 times, good estimates of the variance of these ACFs were produced. Figure 1 plots both Eqs. (6) and (8) along with the simulation results. Since for this case the measurement grid size Δy is small relative to the correlation length ξ, there is very little difference between the discrete and continuous equations. Further, the match to the simulations is sufficiently good that the analytical and simulation curves overlap and are indistinguishable [Fig. 1(a)]. Plotting the difference in variance between simulation and the analytical equations reveals the small differences [Fig. 1(b)]. This difference was scaled to make comparisons more useful, as follows:
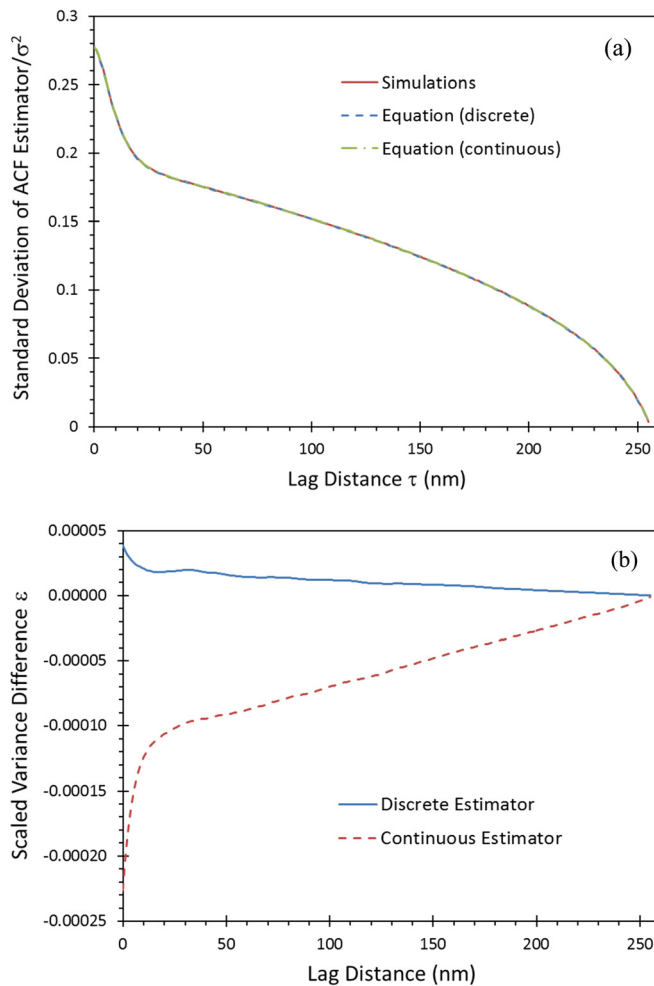

FIG. 1. (Color online) Comparison of the simulated variance (or its square root, the standard deviation) of the estimated ACF to Eqs. (6) and (8) for continuous and discrete cases (biased estimator, $\xi = 10\,nm$, $L = 256\,nm$, $\Delta y = 1\,nm$, and 10^9 simulations): (a) The standard deviation of the estimated ACF showing all three curves on top of each other; (b) the difference in the variance between simulation and equation scaled as in Eq. (9). For these simulations, the generated rough features had $s = 32$ subsampling points for each measurement point to approximate a continuous feature.

$$\text{Continuous: } \varepsilon_c = \frac{T^2}{\sigma^4 L^2}\left(\text{var}(R_c)_{equation} - \text{var}(R_{simulation})\right),$$

$$\text{Discrete: } \varepsilon_d = \frac{K^2}{\sigma^4 N^2}\left(\text{var}(R_d)_{equation} - \text{var}(R_{simulation})\right). \tag{9}$$

As expected and as seen in Fig. 1(b), the discrete Eq. (8) matches the simulation of a discrete ACF better than the continuous Eq. (6).

The simulations suffer from the error of approximating a continuous rough feature with a discretely generated rough feature.[15] The magnitude of the error is a function of the sub-sampling ratio s, the ratio of the measurement grid size Δy to the grid size used to generate the feature. The small values of ε_d shown in Fig. 1(b) are a result of this bias in generating the rough feature using $s = 32$ rather than $s = \infty$.

In most of the work presented below, analysis of a continuous random process will be used to approximate the discrete measurement of random roughness. The difference between Eqs. (6) and (8) quantifies the error in this approach. The largest difference occurs at $m = 0$ (and thus $\tau = 0$). Using a Taylor series expansion for the discrete case

$$\frac{\text{var}(R_d(m=0)) - \text{var}(R_c(\tau=0))}{\text{var}(R_c(\tau=0))} \approx \frac{\Delta y^2}{3\xi^2}\left(\frac{2L+\xi}{2L-\xi}\right). \tag{10}$$

This relative difference is about constant for all τ until τ approaches L for both biased and unbiased estimators. Since in general it will be safe to assume that $L \gg \xi$, the continuous version of the ACF accurately approximates the discrete ACF whenever Δy is much smaller than ξ. For example, there is less than 1% error in estimating the variance using the continuous ACF estimator to approximate the discrete ACF estimator if $\xi > 6\Delta y$ and less than 4% error when $\xi > 3\Delta y$.

It is common advice in both textbooks and research papers that the biased estimator for the ACF is preferred over the unbiased estimator. The reason is clear from Eqs. (6) and (8) since the variance of the unbiased estimate grows as τ approaches L (that is, as T or K approach zero). Figure 2(a) shows that the variance of the biased estimator is always lower than that of the unbiased estimator and goes to 0 as τ approaches L. The downside of using the biased estimator is obviously the bias. The bias is easily determined from Eq. (2) or (3).

$$bias = \langle R_{biased}(\tau)\rangle - \text{ACF}(\tau) = -\frac{|\tau|}{L}\text{ACF}(\tau). \tag{11}$$

The relative bias grows linearly with τ, but since the ACF decreases exponentially with τ, the absolute bias decreases quickly as τ grows. For the important and common case where $L \gg \xi$, the bias remains very small in the region of τ where the ACF is noticeably above 0. As a result, the mean square error (MSE, the sum of the variance of the estimator and the square of its bias) remains lower for the biased estimator than the unbiased estimator. For this reason, the biased
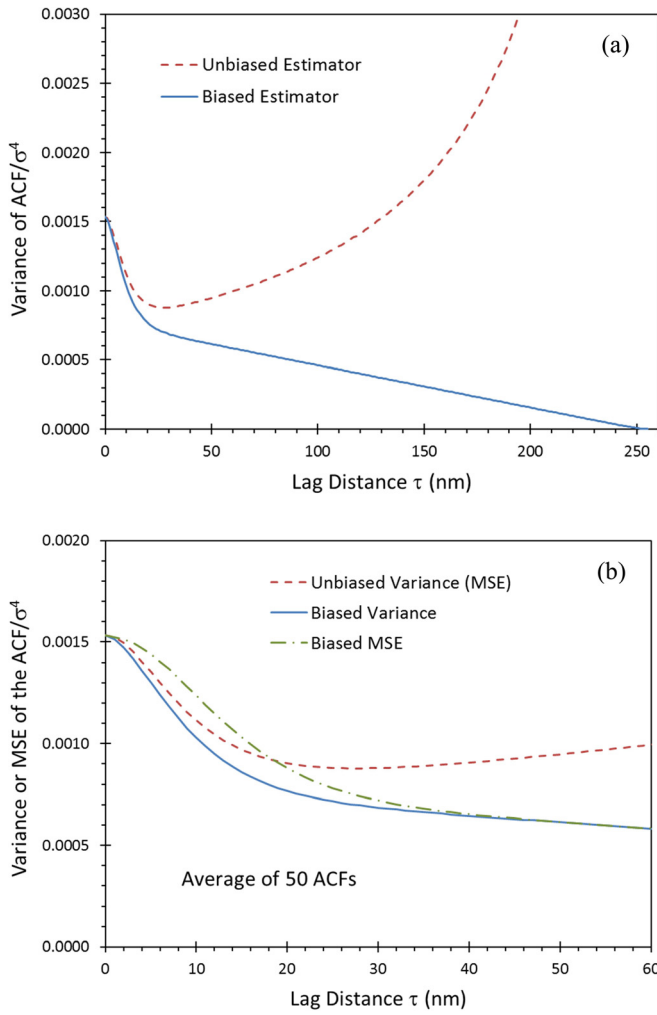
FIG. 2. (Color online) Comparison of biased and unbiased estimators of the ACF ($\xi = 10$ nm, $L = 256$ nm, and $\Delta y = 1$ nm): (a) variance of the estimators from Eq. (6) for $M = 1$; (b) plotting both the variance and the MSE of the estimators for $M = 50$.

estimator of the ACF is the common choice whenever the ACF decays faster than linearly with τ.

For lithography applications, it is common to measure multiple instantiations of the roughness (that is, multiple features that are nominally identical) and average the resulting ACFs. If M features are measured, the average ACF will have a variance reduced by a factor of M. The bias, however, will remain constant. Figure 2(b) shows an example where 50 features have been measured. In this case, the unbiased estimator has a lower MSE for small τ, but higher MSE for larger τ. In general, M must be greater than 20 before the use of an unbiased estimator would improve the low-τ estimates compared to the biased estimates. To improve the estimates for $\tau < \xi$, one must use $M > 20$; to improve the estimates for $\tau < 2\xi$, one must use $M > 60$; and to improve the estimates for $\tau < 3\xi$, one must use $M > 300$. In any case, the use of an unbiased estimator of the ACF is not recommended for $\tau > L/2$.

For the biased ACF estimator, the worst-case variance occurs when $\tau = 0$. For the unbiased ACF estimator, this is also the worst-case variance for $\tau \leq L/2$. Thus, it is useful to

understand how this zero-lag variance scales with the ACF parameters. For the very common case where $L > 3\xi$, Eqs. (6) and (8) can be evaluated at $\tau = m = 0$ to give

$$\frac{\text{var}(R_c(\tau = 0))}{\sigma^4} = \frac{\text{var}(R_c(\tau = 0))}{\langle R_c(\tau = 0) \rangle^2} \approx \frac{2\xi}{L}\left(1 - \frac{\xi}{2L}\right),$$

$$\frac{\text{var}(R_d(m = 0))}{\sigma^4} \approx \frac{2}{N}\left(\frac{1 + \rho^2}{1 - \rho^2}\right)\left\{1 - \frac{2\rho^2}{N(1 - \rho^4)}\right\}$$

$$= \frac{2}{N\tanh(\Delta y/\xi)}\left\{1 - \frac{1}{N\sinh(2\Delta y/\xi)}\right\}. \quad (12)$$

Thus, the variance of the ACF estimate is kept low when $L \gg \xi$. Note also that $R_c(\tau = 0)$ and $R_d(m = 0)$ (and their variances) are the same for the biased and the unbiased estimators.

For the case when $L \gg \xi$ and $\tau \leq L/2$, the variance of the continuous estimator of the ACF can be approximated by

$$\frac{\text{var}(R_c(\tau))}{\sigma^4} \approx \frac{\xi(L - \tau)}{T^2}\left(1 + e^{-2\tau/\xi} + \frac{2\tau}{\xi}e^{-2\tau/\xi}\right), \quad (13)$$

recalling that the unbiased estimator uses $T = L - \tau$ and the biased estimator uses $T = L$. The relative standard deviation of the estimator is the same for the biased and unbiased estimators. For the same case as above, when $L \gg \xi$ and $\tau \leq L/2$,

$$\frac{\sqrt{\text{var}(R_c(\tau))}}{\langle R_c(\tau) \rangle} \approx \sqrt{\frac{\xi}{L - \tau}\left(e^{2\tau/\xi} + 1 + \frac{2\tau}{\xi}\right)}. \quad (14)$$

The relative uncertainty of the ACF estimate grows rapidly as the lag exceeds the correlation length, so that single ACF estimates are not very useful for lags beyond about two correlation lengths.

B. Impact of detrending on ACF estimators

The above analysis assumed a zero mean process (or equivalently, a process where the mean is known exactly). This is never the case for LER or LWR measurements. For LWR, we use the sample mean as an estimate of the true mean, and for LER, we subtract off the best-fit line through the rough edge. The general term for subtracting a sample estimate in place of the population estimate during roughness analysis is *detrending* (also known as subtracting the baseline). Detrending adds considerable systematic bias to the measurement of the ACF.[3,16,17] Figure 3 shows how the measured ACF goes negative and appears to oscillate about zero, even though the true ACF is exponential and always positive.

Detrending also impacts the variance of the ACF estimate. Figure 4 shows the standard deviation of the ACF estimator as well as the square root of the mean square error with and without detrending. While detrending reduces the variance of the estimator [Fig. 4(a)], the bias caused by this detrending adds to the MSE to produce an MSE that is sometimes higher and sometimes lower than the MSE for no
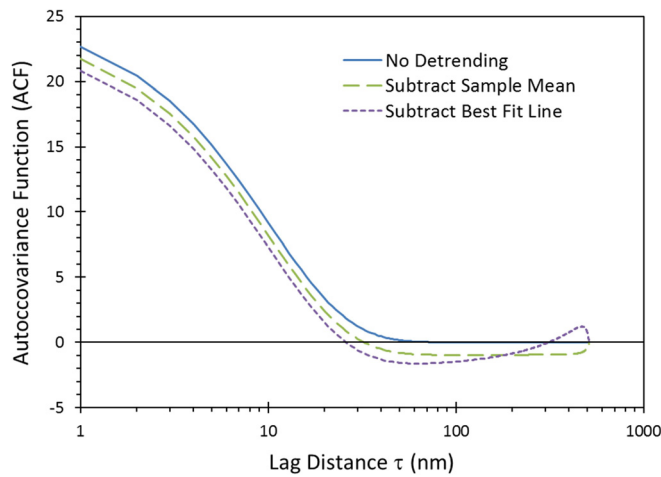
FIG. 3. (Color online) Impact of detrending on the resulting ACF estimate. ACFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, and 10^8 ACFs used to generate these curves. A logarithmic scale for the lag distance was chosen to make the differences in ACF more noticeable. From Ref. 3.

detrending [Fig. 4(b)]. For the most important part of the ACF (moderate to small lags), mean detrending (for LWR) produces a smaller MSE than no detrending, and linear detrending (for LER) produces about the same MSE as no detrending (assuming $M = 1$, only a single feature is measured).

C. Impact of measurement averaging on ACF estimators

So far, our simulation of the measurement process assumed a measurement probe that can precisely measure the width/edge position of the rough feature at the exact y-coordinate of the measurement. Practical measurement tools, such as an atomic force microscope or a scanning electron microscope (SEM), have a probe size greater than zero, so that the measurement result is a convolution of the probe response and the rough edge. For example, the width/edge position as measured by an SEM can be approximated as the weighted average of the rough edge multiplied by a Gaussian electron beam shape centered at the measurement position. This spatial averaging effect has been shown to reduce the apparent high-frequency roughness when extracting the power spectral density.[2] Note also that SEM image analysis can include the application of filters, such as a Gaussian filter, before edge detection that can result in some of the same effects as spatial measurement averaging. Those
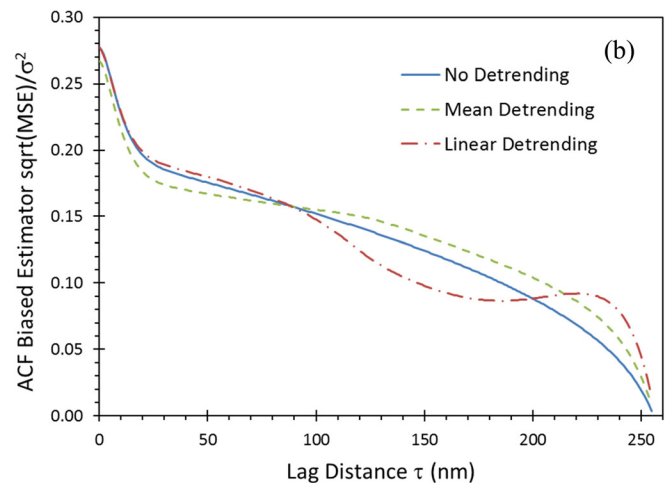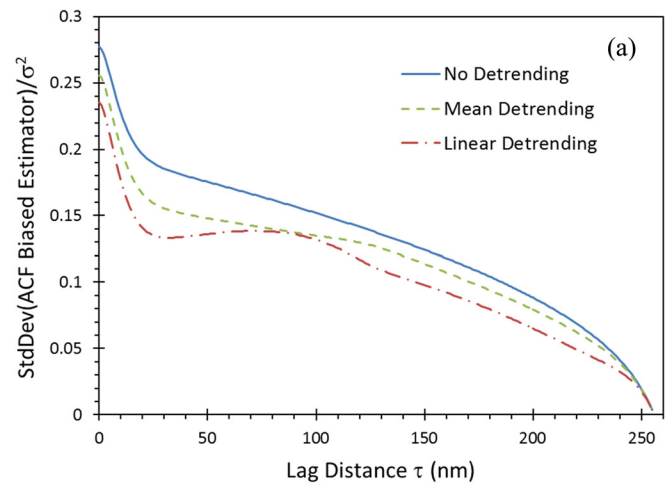

FIG. 4. (Color online) Impact of detrending on the variance of the biased ACF estimator: (a) square root of the variance of the ACF estimator, and (b) square root of the MSE of the ACF estimator. ACFs were generated using simulation with $M = 1$, $N = 256$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, and 10^8 ACFs used to generate these curves.

effects have been considered elsewhere[18,19] but will not be considered here.

As the results below show, spatial measurement averaging has a corresponding bias on the estimated ACF for small lag τ. Figure 5 shows that the bias caused by measurement averaging reduces the estimated ACF for lag distances on the order of the Gaussian FWHM probe size or less. Further, simulations show that the variance of the estimated ACF is mostly independent of the measurement probe size. The magnitude of the zero lag bias is found empirically to be about

$$\frac{\text{ACF}_d(\tau = 0, \, probe\,width = 0) - \text{ACF}_d(\tau = 0, \, probe\,width = pw)}{\sigma^2} \approx \frac{pw}{2\xi}, \tag{15}$$

where pw is the Gaussian probe FWHM. This bias due to measurement averaging is a strong function of the shape

distribution used for the averaging. If instead of a Gaussian probe shape a Lorentzian shape is assumed (to account for
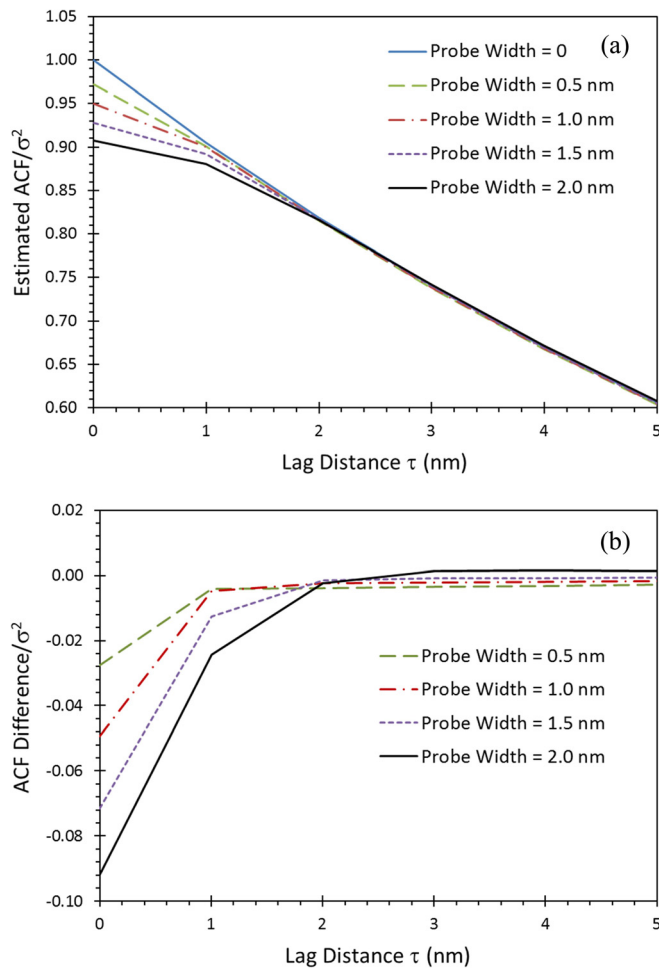
FIG. 5. (Color online) Impact of spatial averaging during measurement on the ACF: (a) the simulated ACF as a function of probe width (FWHM), and (b) the difference between the ACF with measurement averaging and the ACF for a probe width of zero. ACFs were generated using simulation with $N = 256$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $s = 32$, and 10^7 ACFs averaged to generate these curves.

long-range backscatter electrons), the impact of averaging is about five times greater for the same FWHM owing to the very heavy tails of the Lorentzian distribution.

D. Impact of measurement noise on ACF estimators

Another complication in LER/LWR measurement is noise in the measurement itself. For example, SEM images are made of pixels with grayscale values determined by the counts of secondary electrons detected when the electron beam impinges on that point in the sample. In many cases, the electron dose must be kept low to prevent modification of the sample, so that a relatively small number of secondary electrons are counted per pixel. This gives rise to shot noise following a Poisson distribution so that the variance in the electron count of a pixel equals the mean of the electron count. This shot noise creates the "snowy" appearance so common in SEM images and contributes to random errors in the detected edge position. These errors are nearly uncorrelated, and the effect can be approximated as a white-noise error (ignoring the resolution limit of the SEM imaging optics) added to the actual feature roughness.[20]

From the perspective of the ACF, adding uncorrelated (white) noise to the measured data will cause only an increase in the zero-lag ACF (all other lag values being unaffected)

$$\text{ACF}(\tau = 0) = \sigma_{apparent}^2 = \sigma_{LWR}^2 + \sigma_{MN}^2, \tag{16}$$

where σ_{LWR}^2 is the true LWR variance and σ_{MN}^2 is the variance of the measured linewidth due to metrology noise. Interestingly, metrology noise is of similar magnitude but works in the opposite direction as measurement averaging due to nonzero measurement probe width. While metrology probe averaging lowers $\sigma_{apparent}^2$ (see Fig. 5), metrology noise increases it. This will be discussed further in the context of the height–height covariance function in Sec. III.

III. HHCF UNCERTAINTY

The height–height covariance function is less commonly used than the autocovariance function for describing correlated roughness behavior, but it has some distinct advantages. The HHCF is defined as

$$\text{HHCF}(\tau) = \left\langle (w(s) - w(s + \tau))^2 \right\rangle. \tag{17}$$

[Note that Constantoudis et al.[8–10] have used a somewhat less common definition of the HHCF, equal to the square root of Eq. (17).] By carrying out the square in this definition, it is easy to relate the HHCF to the ACF:

$$\text{HHCF}(\tau) = 2(\sigma^2 - \text{ACF}(\tau)) = 2(\text{ACF}(0) - \text{ACF}(\tau)). \tag{18}$$

In other fields, the HHCF is referred to as the structure function or sometimes the variance function.[21]

The most important advantage of the HHCF is that detrending by subtracting the mean has no impact on the HHCF, since it is computed from the difference between measurements. Thus, for LWR measurement, the HHCF is unbiased. For a finite line length, the standard discrete and continuous unbiased estimators for the HHCF are given by, respectively,

$$\text{HHCF}_d(\tau) = \frac{1}{N-m} \sum_{n=0}^{N-1-m} \left(w(s) - w(s + \tau) \right)^2,$$
$$\text{HHCF}_c(\tau) = \frac{1}{L-\tau} \int_0^{L-\tau} \left(w(s) - w(s + \tau) \right)^2 ds. \tag{19}$$

A. Variance of HHCF estimators

The variance of each of these estimators can be worked out by first carrying out the squares in Eq. (19). For the continuous estimator, this leads to

$$\text{HHCF}_c(\tau) = w^2(s)_{avg} + w^2(s + \tau)_{avg} - 2R_c(\tau), \tag{20}$$

where $w^2(s)_{avg} = (1/L - \tau) \int_0^{L-\tau} w^2(s) ds$, $w^2(s + \tau)_{avg} = (1/L - \tau) \int_0^{L-\tau} w^2(s + \tau) ds$.

Calculating the variance of Eq. (20) produces several covariance terms

$$
\mathrm{var}(\mathrm{HHCF}_c(\tau)) = \mathrm{var}(w^2(s)_{avg}) + \mathrm{var}(w^2(s+\tau)_{avg}) + 4\mathrm{var}(R_c(\tau))
$$
$$
+ 2\mathrm{cov}(w^2(s)_{avg}, w^2(s+\tau)_{avg}) - 4\mathrm{cov}(w^2(s)_{avg}, R_c(\tau)) - 4\mathrm{cov}(w^2(s+\tau)_{avg}, R_c(\tau)). \tag{21}
$$

The variance of the mean square value of the data is equal to the variance of $R_c(0)$ for a line length of $L - \tau$. Consider now the case of the exponential autocovariance function given by Eq. (5). Following the same basic procedure used by Fuller[14]

$$
\frac{\mathrm{var}\left(w^2(s)_{avg}\right)}{\sigma^4} = \frac{\mathrm{var}\left(w^2(s+\tau)_{avg}\right)}{\sigma^4} = \frac{\xi^2}{(L-\tau)^2}\left\{ e^{-2(L-\tau)/\xi} + \frac{2(L-\tau)}{\xi} - 1 \right\}. \tag{22}
$$

The first covariance term in Eq. (21) becomes

$$
\frac{\mathrm{cov}\left(w^2(s)_{avg}, w^2(s+\tau)_{avg}\right)}{\sigma^4} = \frac{\xi^2}{(L-\tau)^2}\left\{ e^{-2(L-\tau)/\xi}\cosh\left(\frac{2\tau}{\xi}\right) - e^{-2\tau/\xi} + \frac{2(L-2\tau)}{\xi} \right\} \quad \text{for } 0 \le \tau \le L/2
$$
$$
= \frac{\xi^2}{(L-\tau)^2} e^{-2\tau/\xi}\left\{ \cosh\left(\frac{2(L-\tau)}{\xi}\right) - 1 \right\} \quad \text{for } L/2 \le \tau \le L. \tag{23}
$$

The final two covariance terms are equal to each other

$$
\frac{\mathrm{cov}\left(w^2(s)_{avg}, R_c(\tau)\right)}{\sigma^4} = \frac{\xi^2}{(L-\tau)^2}\left\{ e^{-2(L-\tau)/\xi}\cosh\left(\frac{\tau}{\xi}\right) + e^{-\tau/\xi}\left[\left(\frac{2(L-\tau)}{\xi} - \frac{\tau}{\xi}\right)\left(1+\frac{\tau}{\xi}\right) - 1 \right] \right\} \quad \text{for } 0 \le \tau \le L/2
$$
$$
= \frac{\xi^2}{2(L-\tau)^2} e^{-\tau/\xi}\left\{ e^{-2(L-\tau)/\xi} + \frac{2(L-\tau)}{\xi} - 1 + \frac{2(L-\tau)^2}{\xi^2} \right\} \quad \text{for } L/2 \le \tau \le L. \tag{24}
$$

Combining all terms gives the final result, an example of which is shown in Fig. 6.

A useful approximation to the result for the case where $\xi \ll L$ is

$$
\frac{\mathrm{var}(\mathrm{HHCF}_c(\tau))}{\sigma^4} \approx \frac{4\xi}{(L-\tau)}\left\{ 3 + \left(1+\frac{2\tau}{\xi}\right)e^{-2\tau/\xi} - 4\left(1+\frac{\tau}{\xi}\right)e^{-\tau/\xi} \right.
$$
$$
\left. + \frac{\tau}{(L-\tau)}\left[\left(1+\frac{\tau}{\xi}\right)\left(2e^{-\tau/\xi} - e^{-2\tau/\xi}\right) - 1 \right] \right\} \quad \text{for } 0 \le \tau \le L/2
$$
$$
\approx \frac{4\xi^2}{(L-\tau)^2}\left\{ e^{-2(L-\tau)/\xi} + \frac{2(L-\tau)}{\xi} - 1 \right\} \quad \text{for } L/2 \le \tau \le L. \tag{25}
$$

Another useful approximation, for the relative standard deviation of the HHCF, is valid for $3\xi < \tau \le L/2$

$$
\frac{\sqrt{\mathrm{var}(\mathrm{HHCF}_c(\tau))}}{\mathrm{HHCF}_c(\tau)} \approx \sqrt{\frac{3\xi}{L}}\left(1 + \frac{\tau}{3L}\right). \tag{26}
$$

For values of $\tau < 3\xi$, the relative standard deviation decreases to zero at τ approaches 0. Thus, the highest accuracy estimates for the HHCF occur for small values of the lag distance τ. This makes the HHCF a useful tool for determining both the correlation length and the roughness exponent. The uncertainty in the HHCF grows rapidly for $\tau > L/2$, so that the HHCF in this region should probably be ignored.

It should be noted that a biased estimator similar to that used for the ACF is not appropriate for the HHCF. The use of a biased estimator for the ACF works well when the ACF declines to zero faster than linearly with τ. Since the HHCF does not go to zero for large τ (it approaches $2\sigma^2$), a biased estimator like that used for the ACF would unacceptably distort the HHCF.

B. Impact of detrending on HHCF estimators

As already mentioned, one of the benefits of using the HHCF over the ACF is that the HHCF is not biased when detrending the sample mean. It is, however, biased when subtracting the best-fit line, as is commonly done for LER measurement. Figure 7 shows that the HHCF is biased lower for large values of τ due to linear detrending. In this case, a bias of about 10% occurs at
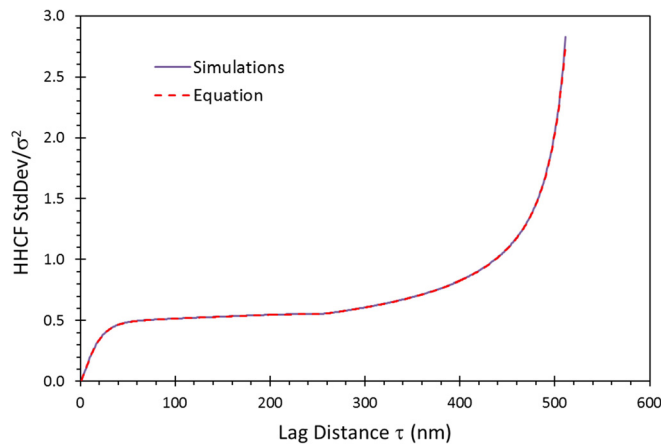
FIG. 6. (Color online) Square root of the variance of the continuous HHCF using the derived equations and as generated by simulations ($N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, and 10^7 HHCFs used to generate the simulation curve).

$\tau = L/2$, and the fractional bias is about equal to $3\xi/L$ at this lag. Fortunately, the worst bias occurs in a region of the lag distance that would generally be avoided (that is, $\tau > L/2$). The linear detrending also lowers the variance of the HHCF, but the MSE remains about the same with or without linear detrending for $\tau < L/2$.

C. Impact of measurement averaging on HHCF estimators

As seen before, a nonzero measurement probe width causes averaging that smoothens some of the high-frequency roughness. For the ACF, that resulted in a drop in the ACF for values of τ near 0. For the HHCF, the impact of measurement probe averaging is a near-constant drop in the HHCF for all values of τ. Figure 8 shows a typical example of the impact of measurement averaging on the HHCF. Figure 8(a) shows the HHCF, while Fig. 8(b) shows the difference between the HCF with measurement averaging and the
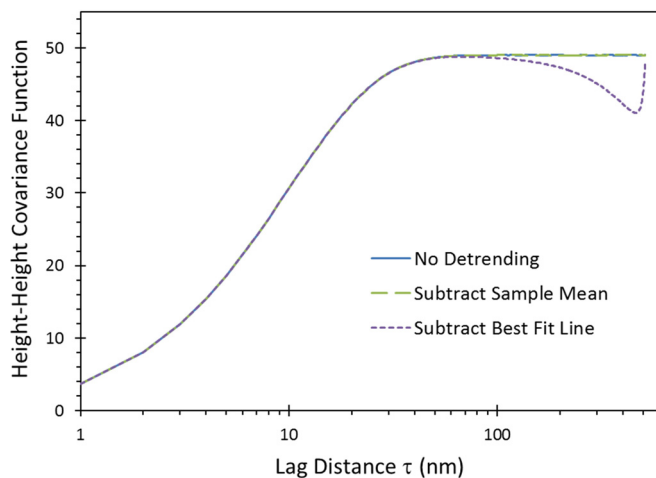

FIG. 7. (Color online) Impact of linear detrending on the resulting HHCF. HHCFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, and $M = 10^7$ HHCFs averaged together. A logarithmic scale for the lag distance was chosen to make the differences in HHCF more noticeable. From Ref. 3.
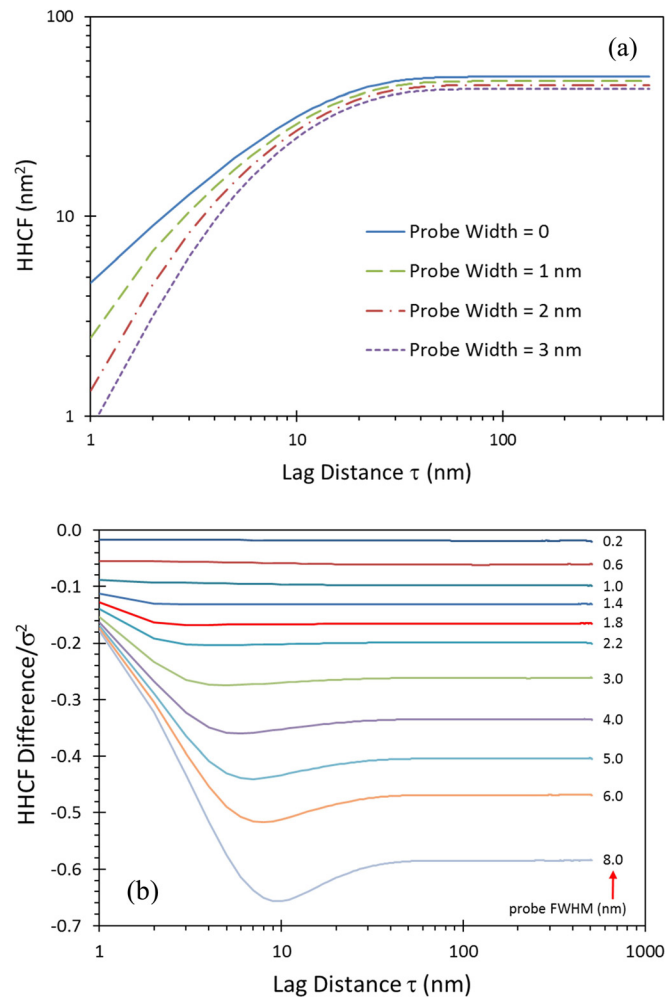

FIG. 8. (Color online) Impact of measurement probe averaging on the resulting HHCF: (a) the HHCF for different probe widths (FWHM), and (b) the difference between the HHCF with measurement averaging and the HHCF for a probe width of zero, normalized by dividing by the true feature variance. The probe FWHM (nm) is shown by the label next to each curve. HHCFs were generated using simulation with $N = 512$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, subsampling $s = 16$, and $M = 10^7$ HHCFs averaged together.

HHCF for a probe width of zero (no averaging), normalized by dividing by the true feature variance (σ^2).

For moderate probe widths, the impact of averaging is a uniform decrease in the HHCF. The magnitude of the decrease is given approximately by

$$\frac{\mathrm{HHCF}_d(probe\ width = 0) - \mathrm{HHCF}_d(probe\ width = pw)}{\sigma^2}$$
$$\approx \frac{pw}{\xi}, \tag{27}$$

where pw is the FWHM of the Gaussian averaging function. Thus, the impact of measurement averaging can be minimized by keeping the probe width small compared to the correlation length.

D. Impact of measurement noise on HHCF estimators

As mentioned above and shown in Eq. (16), the impact of measurement noise is a simple increase in the measured
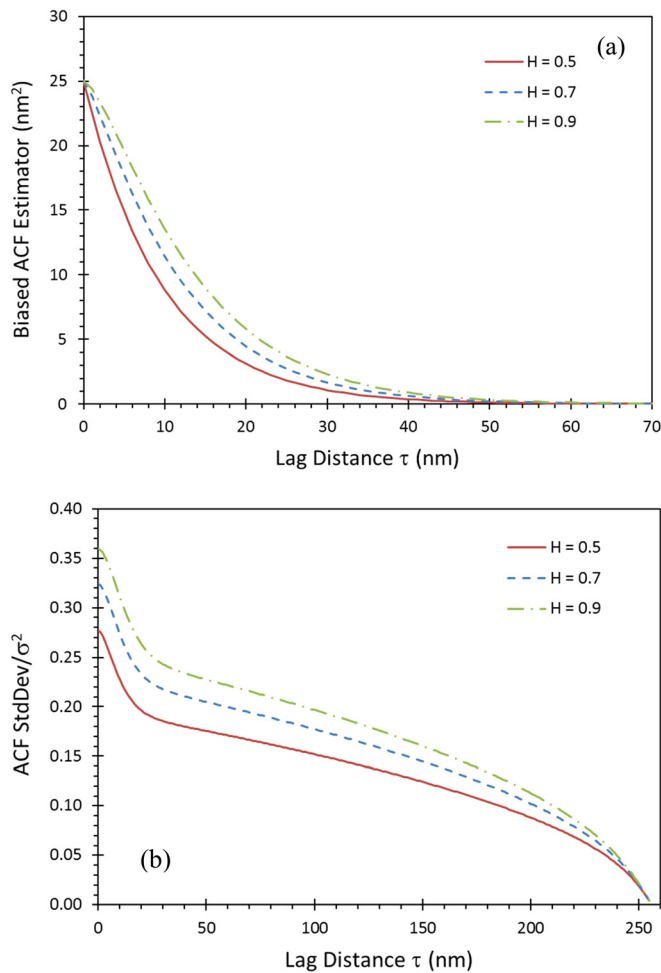
FIG. 9. (Color online) Impact of the Hurst roughness exponent on the simulated ACF: (a) the biased ACF, and (b) the standard deviation of the estimated ACF. Simulations used $N = 256$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, subsampling $s = 8$, and 10^8 simulations.
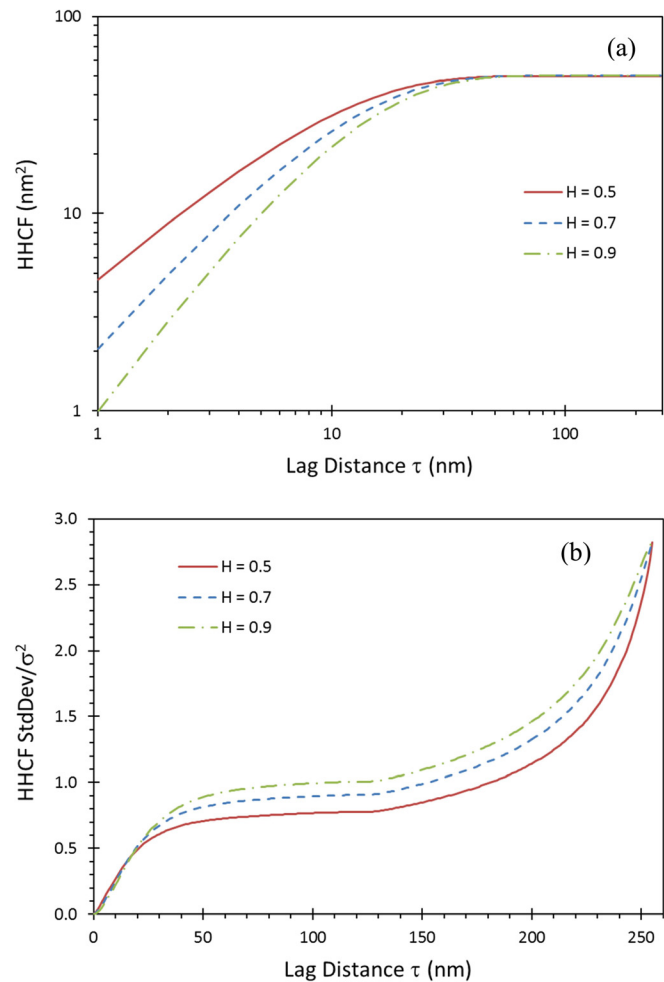
FIG. 10. (Color online) Impact of the Hurst roughness exponent on the simulated HHCF: (a) the estimated HHCF, and (b) the standard deviation of the estimated HHCF. Simulations used $N = 256$ points, $\Delta y = 1$ nm, $\xi = 10$ nm, $\sigma = 5$ nm, subsampling $s = 8$, and 10^8 simulations.

(apparent) variance of the roughness. For the case of the HHCF, this results in a uniform increase in the measured HHCF by an amount $2\sigma^2_{MN}$. Simulations bear out this expected result. Note that, as was the case for the ACF,

measurement noise works oppositely to measurement probe averaging. Small amounts of measurement probe averaging could be compensated by a small amount of measurement noise under the right circumstances. Comparing Eqs. (16)

TABLE I. Summary of issues affecting the estimation of the PSD, the ACF, and the HHCF for typical roughness data.

Measurement condition	PSD	ACF	HHCF
$\Delta y > 0$	Aliasing, PSD is too high for high frequencies	No impact on expected value; discrete estimator variance grows as Δy approaches ξ	No impact on expected value; discrete estimator variance grows as Δy approaches ξ
Finite L	Leakage, PSD is too low for low frequencies	No impact on expected value; variance goes as ξ/L	No impact on expected value; variance goes as ξ/L
Random variance	Standard deviation of PSD = mean of PSD for each frequency	Greatest variance at $\tau = 0$; unbiased estimator should not be used for $\tau > L/2$	Grows very large for $\tau > L/2$; HHCF should not be used for $\tau > L/2$
Mean detrending	Impossible to determine PSD(0)	Biases ACF for all τ, makes ACF go negative	No impact
Linear detrending	Reduces very low-frequency PSD	Biases ACF for all τ, makes ACF go negative	Biases HHCF for large τ
Measurement probe averaging	Reduces high-frequency PSD	Reduces ACF for τ near 0	Reduces HHCF for all τ
Metrology noise	Increases PSD for all frequencies	Increases ACF for $\tau = 0$	Increases HHCF for all τ

and (27) for a Gaussian probe shape, such compensation occurs when

$$\sigma_{MN} \approx \sigma_{LWR} \sqrt{\frac{pw}{2\xi}} \quad \text{or} \quad pw \approx 2\xi \left(\frac{\sigma_{MN}}{\sigma_{LWR}} \right)^2 . \quad (28)$$

IV. INFLUENCE OF THE HURST ROUGHNESS EXPONENT

In the discussion so far, a specific form of the ACF was chosen: the exponential of Eq. (5). While common, this is not the only autocovariance function behavior found in roughness data from lithographically printed resist or after-etch features. For example, most LER/LWR data have been found to be well described by the Palasantzas power spectral density function[22]

$$\text{PSD}(f) = \frac{\text{PSD}(0)}{\left[1 + (2\pi f \xi)^2 \right]^{H+1/2}}, \quad (29)$$

where H is the Hurst roughness exponent, and PSD(0) is found (using the gamma function, Γ) by

$$\text{PSD}(0) = 2\sigma^2 \xi \left(\frac{\sqrt{\pi} \, \Gamma \left(H + \frac{1}{2} \right)}{\Gamma(H)} \right). \quad (30)$$

The PSD and the ACF form a Fourier transform pair. When $H = 0.5$, the Palasantzas PSD is the Fourier transform of the exponential ACF of Eq. (5). For other values of H, the analytical derivations given in Secs. II and III of this paper do not strictly apply. However, simulation can still be used to investigate the variance of the ACF and HHCF estimates for $H > 0.5$.

Figure 9 shows the simulated ACF (using the biased discrete estimator) and the standard deviation of the ACF. The general shape of the ACF standard deviation curve is the same as a function of roughness exponent H. Roughly speaking, the $H = 0.7$ case has 17% larger standard deviation than the $H = 0.5$ case, and the $H = 0.9$ case has about 30% greater standard deviation than $H = 0.5$. Figure 10 shows the simulated HHCF and its standard deviation for different values of roughness exponent H. For lag distances of about twice the correlation length or less, increasing H reduces the mean value of the HHCF but has little impact on its uncertainty. For longer lag distances, higher H does not impact the mean HHCF but causes a greater standard deviation of the HHCF estimate.

V. CONCLUSIONS

The roughness of a feature edge or width is commonly characterized by the amount of roughness, the standard deviation of the difference between the actual edge or width and its target. But the presence of correlations between different points along the line edge requires a more thorough analysis. The most common tools for characterizing roughness are the PSD, the ACF, and the HHCF. These three functions are related to each other, so that perfect knowledge of one determines the other two.

Since perfect knowledge of any of these functions is not possible, they must be estimated from experimental data. In previous studies, the uncertainties and biases associated with nonparametric estimation of the PSD were described.[2,3]
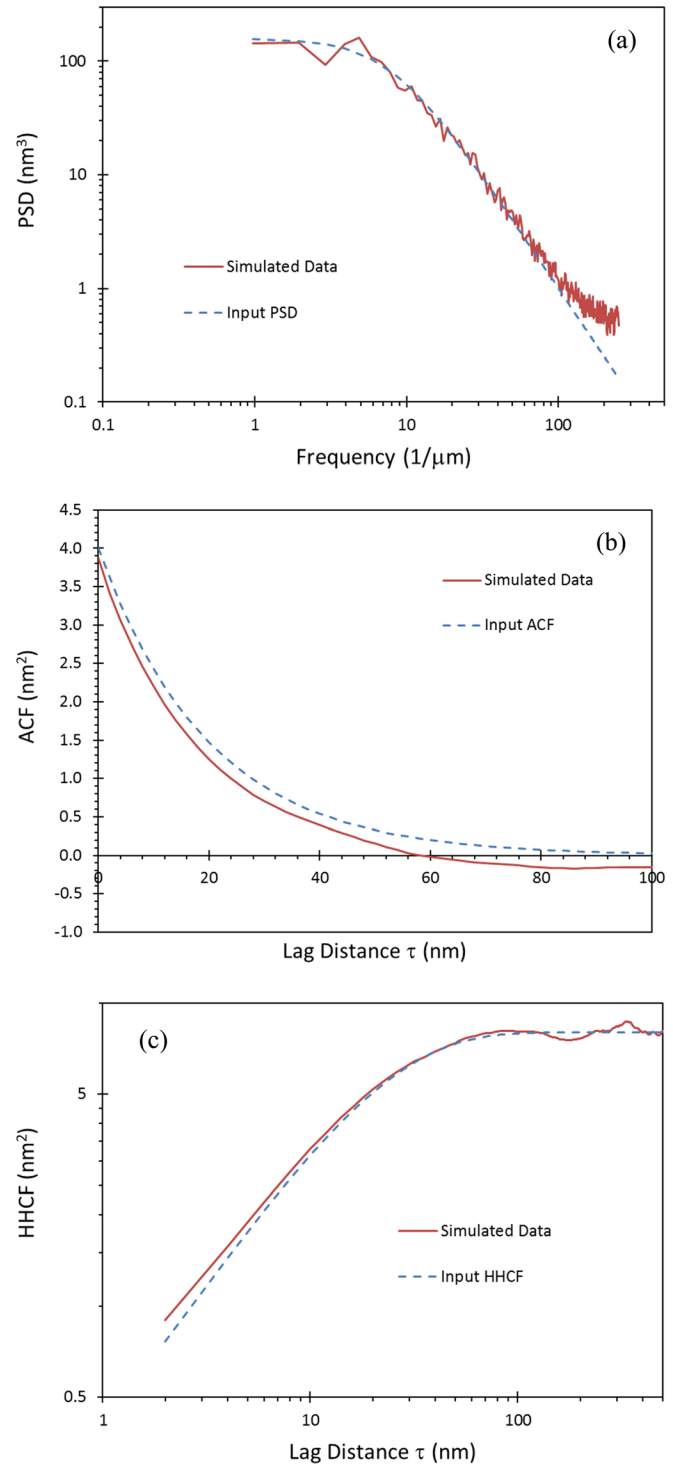

FIG. 11. (Color online) Comparisons of roughness analysis approaches for simulated rough features: (a) the estimated PSD compared to the simulation input, (b) the estimated ACF compared to the simulation input, and (c) the estimated HHCF compared to the simulation input. Simulations used $N = 512$ points, $\Delta y = 2 \, \text{nm}$, $\xi = 20 \, \text{nm}$, $\sigma = 2 \, \text{nm}$, subsampling $s = 32$, $H = 0.5$, Gaussian metrology probe FWHM = 2 nm, metrology noise = 0.5 nm, mean detrending, Welch taper window for PSD calculation, and $M = 50$ simulated rough lines used for averaging.

In this work, the uncertainties and biases associated with nonparametric estimation of the ACF and the HHCF are discussed for the case of the commonly encountered exponential autocorrelation function.

Table I summarizes what has been learned about nonparametric estimation of roughness characterization functions. Unfortunately, there is no one best function for characterizing roughness from experimental data. The PSD suffers from leakage and aliasing. One advantage of using the ACF or HHCF over the PSD is that they do not. The ACF and HHCF, however, suffer from other complications.

For the ACF, the biased estimator is almost always preferred due to its lower variance. Still, the variance is sufficiently large that little or no information can be obtained for lag distances beyond a few correlation lengths. Further, the ACF is biased due to detrending, making the ACF a poor replacement for the PSD. The HHCF has the advantage that it is not biased due to mean detrending, as would be done when measuring LWR. Unlike the ACF, it has a nearly constant relative uncertainty out to lag distances of half the measured feature length, and an uncertainty that goes to zero as τ decreases below a few correlation lengths. Thus, the HHCF is a viable alternative to the use of the PSD for characterizing the frequency (lag) behavior of a feature's roughness.

As a final comparison, all of the effects described in this paper have been combined in a simulation of a typical rough linewidth measurement scenario. For 50 measurements averaged together, the resulting PSD, ACF, and HHCF are shown in Fig. 11. For the PSD, the root-mean-square-error is about 15% at low frequencies, but grows to 75% at the highest frequencies due to aliasing and metrology noise. For the ACF, the root-mean-square-error begins low at 4% for $\tau = 0$, but reaches 10% by one correlation length, 27% by three correlation lengths, and 100% by five correlation lengths. For the HHCF, the largest relative root-mean-square-error is at a lag of one grid point, 15%. By one correlation length the root-mean-square-error has settled down to the 3%–4% level, where it remains through $\tau < L/2$. Thus, under this scenario, the HHCF is the lowest-error approach to roughness characterization.

There are two fundamental kinds of length scales at work here. The physical length scale is the correlation length, ξ, so that all other physical lengths can be thought of as multiples of the correlation length. But there are also measurement length scales: the probe width and the sampling distance. Measurements have lengths that can be described as multiples of the sampling distance and/or probe width. When measuring a physical process that follows the statistics assumed to apply here, the outcomes contain a mixture of these two sets of length scales. The mathematical results derived in this paper help to understand the combined role of these length scales in PSD, ACF, and HHCF measurements.

All of the estimation approaches studied here fall under the category of nonparametric estimation, meaning that the estimators used do not require any *a priori* knowledge of the expected shape of the function to be estimated. Future work should explore the use of parametric estimators, where an expected model form, such as the Palasantzas PSD, is used to improve the estimate, lowering its uncertainty.

[1]C. A. Mack, J. Micro/Nanolithogr. MEMS MOEMS **13**, 020501 (2014).
[2]C. A. Mack, J. Micro/Nanolithogr. MEMS MOEMS **12**, 033016 (2013).
[3]C. A. Mack, J. Micro/Nanolithogr. MEMS MOEMS **14**, 033502 (2015).
[4]T. Verduin, P. Kruit, and C. W. Hagen, J. Micro/Nanolithogr. MEMS MOEMS **13**, 033009 (2014).
[5]A. Hiraiwa and A. Nishida, J. Micro/Nanolithogr. MEMS MOEMS **9**, 041210 (2010).
[6]L. Azarnouche, E. Pargon, K. Menguelti, M. Fouchier, M. Brihoum, R. Ramos, O. Joubert, P. Gouraud, and C. Verove, J. Micro/Nanolithogr. MEMS MOEMS **12**, 041304 (2013).
[7]A. Vaglio Pret, R. Gronheid, T. Ishimoto, and K. Sekiguchi, J. Micro/Nanolithogr. MEMS MOEMS **9**, 041308 (2010).
[8]V. Constantoudis, G. P. Patsis, A. Tserepi, and E. Gogolides, J. Vac. Sci. Technol., B **21**, 1019 (2003).
[9]V. Constantoudis, G. P. Patsis, L. H. A. Leunissen, and E. Gogolides, J. Vac. Sci. Technol., B **22**, 1974 (2004).
[10]V. Constantoudis, G. P. Patsis, and E. Gogolides, J. Micro/Nanolithogr. MEMS MOEMS **3**, 429 (2004).
[11]P. P. Naulleau and J. P. Cain, J. Vac. Sci. Technol., B **25**, 1647 (2007).
[12]M. S. Bartlett, Suppl. J. R. Stat. Soc. **8**, 27 (1946).
[13]G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, 1968), pp. 177–178.
[14]A. T. Fuller, J. Electr. Control **4**, 551 (1958).
[15]C. A. Mack, Appl. Opt. **52**, 1472 (2013).
[16]E. L. O'Neill and A. Walther, J. Opt. Soc. Am. **67**, 1125 (1977).
[17]E. R. Freniere, E. L. O'Neill, and A. Walther, J. Opt. Soc. Am. **69**, 634 (1979).
[18]G. P. Patsis, V. Constantoudis, A. Tserepi, E. Gogolides, and G. Grozev, J. Vac. Sci. Technol., B **21**, 1008 (2003).
[19]A. Hiraiwa and A. Nishida, J. Micro/Nanolithogr. MEMS MOEMS **10**, 23010 (2011).
[20]J. S. Villarrubia and B. D. Bunday, Proc. SPIE **5752**, 480 (2005).
[21]R. S. Sayles and T. R. Thomas, Wear **42**, 263 (1977).
[22]G. Palasantzas, Phys. Rev. B **48**, 14472 (1993).