Monte-Carlo Evaluation of Bias and Variance in Hurst Exponents Computed from Power Spectral Analysis of Atomic Force Microscopy Topographic Images

Robert Chrostowski[1,2], Zixuan Li[1,2], James Smith[3], and Filippo Mangolini[1,3*]
1. Texas Materials Institute, The University of Texas at Austin, Austin, Texas 78712, USA
2. Materials Science and Engineering Program, The University of Texas at Austin, Austin, Texas 78712, USA and
3. Walker Department of Mechanical Engineering,
The University of Texas at Austin, Austin, Texas 78712, USA
(Dated: August 4, 2021)

Surface topography influences several surface properties, including friction and adhesion. While a statistical description of surface topography can be obtained from a power spectral density (PSD) analysis of atomic force microscopy (AFM) height maps and fitting the self-affine region of the PSD to determine the Hurst exponent (H), the accuracy of this approach has not been evaluated yet. Here, we use a Fourier filtering algorithm combined with a novel approach to simulate typical AFM scan-line anisotropy to generate synthetic AFM topography images with *known* input Hurst exponent. A Monte Carlo approach is then used to evaluate the variance and *bias* in H estimation from PSDs across different hypothetical experimental approaches, including the case of a cluster of images collected at one scan size (scale) and the case of a cluster of images collected at different scales. Our analysis reveals that estimates of the Hurst exponent from images collected at a single scale are persistently biased in a scale-dependent fashion despite misleading convergence in variance. This bias can be reduced by combining images collected at least at three different scales across the range of scales accessible to AFM.

I. INTRODUCTION

At sufficiently fine scales, all surfaces exhibit texture, which can dramatically affect adhesive [1–3], optical [4], thermal [5, 6] and mechanical [7] properties of materials [8, 9]. Formal surface finish standards (e.g. ASME B46.1, ISO 4287, ISO 4288, ISO 25178) enumerate many ways to parameterize both small- (roughness) and large- (waviness) scale texture, and different parameterizations are relevant for quantifying the impact on different properties [10, 11]. For example, contact stiffness of an interface depends on the root-mean-square (RMS) height [12–15] whereas adhesion depends on RMS curvature [2]. Complex wettability modeling [16, 17], which has been instrumental to engineering both classic lotus leaf-type superhydrophobic surfaces [18] as well as more recent lubricant-infused surfaces [19], depends strongly on the ratio of a surface's true area to its projected area (r). This area ratio also impacts thermal transport properties [20] and contact mechanics [21, 22]. Accurately modeling the quantitative relationships between these important properties and the underlying area ratio is impeded by two critical challenges:

Challenge 1: Most laboratory experiments seeking to study these relationships are performed on patterned surfaces [4–6, 23–25] for which the area ratio can be computed to sufficient accuracy using simple geometric formulae, whereas most surfaces in practical engineering applications are machined and thus stochastic.

Challenge 2: Obtained values for the area ratio on stochastic surfaces can vary by two orders of magnitude or more [11] depending on the scale of measurement. This is analogous to Mandlebrot's observations with respect to finding the "true" length of a coastline [26, 27].

Stochastic surfaces can be effectively described across many scales by a power spectral density (PSD or ϕ). The PSD has two equivalent definitions as either the Fourier transform of the autocovariance function (the correlogram definition), or as the expectation value of square modulus of the normalized Fourier transform (the periodogram definition) [28]. These can be computed directly from line profiles (SEMI MF1811) or extended to two-dimensional surface maps as we carry out in the present study (see Jacobs et al. [10] for a comprehensive review of the different curves that are called "PSD"). In either case, the PSD contains amplitude information as a function of frequency, but discards phase information. The use of *power* in the name originates in electrical engineering, where, if the series under study is voltage as a function of time, by virtue of Parseval's theorem the value of $\phi(\omega)$ represents the contribution at any given frequency to the overall power of the series [29]. For analysis of surfaces, the PSD might be better named the *Height Variance Spectrum*, as the value of $\phi(k_x, k_y)$ represents the contribution of each *spatial* frequency (as angular wavenumbers k in [radians/m]) to the square of the RMS height, i.e., the variance of $h(x, y)$ about a mean plane $h = 0$. The primary utility of this description for surfaces is that Parseval's theorem contains the RMS height, slope and curvature as simple sum rules [10]. Both computational simulations [30] and experimental evidence [11] suggest that the PSD continues to have physically meaningful values down to absolute smallest scales (i.e., atomic spacing). Moreover, due to the mathematical definition of Parseval's theorem for derivatives, the "true" quantitative values for RMS slope (from which the area ratio is

* filippo.mangolini@austin.utexas.edu

derived) and RMS curvature of a surface are dominated by these smallest scales [9, 10] (see the Supplemental Material for a more in-depth discussion of the underlying mathematics).

Even though roughness at the smallest scales can be directly investigated through different experimental methods (e.g., transition electron microscopy (TEM) cross-sections [11]), this is impractical for large-scale or routine measurements. The decrease in costs and increase in efficiency with the wide-scale adoption of atomic force microscopy (AFM) [31, 32], coupled with its ability to image scales smaller than those of diffraction-limited optical profilometry [10, 33, 34], make it an ideal tool for practical investigations. However, the limited lateral resolution of AFM [32], particularly when imaging rough surfaces, does not enable accessing topography scales smaller than the radius of curvature of the tip or about $5 - 10$ $[nm]$. This presents a challenge to answering questions about parameters such as RMS slope or RMS curvature which are dominated by the topography at the smallest scales. This challenge can partially be addressed by modeling and extrapolating the available data to the relevant scales beyond the lateral resolution of the AFM. For the finer scales of many engineering surfaces a fractal or self-affine model is appropriate[35, 36], in which case the PSD can be modeled with a two parameter power law consisting of the pre-factor (C_0) and tail exponent (H). The later parameter, called the Hurst exponent (related to the fractal dimension D by $D = 3 - H$) [27, 37], is critical for the accurate extrapolation to the finest length scales. Across many fields [11, 38–40] the value of the Hurst exponent is commonly estimated through ordinary-least-squares (OLS) regression of a doubly logarithmic transformation of the PSD. However, the physical nature of AFM experiments and generation of AFM topographic images pose several challenges, which can introduce substantial bias to estimates of the Hurst exponent obtained using doubly logarithmic OLS:

Challenge 1: AFM has a slow and a fast scan direction, which introduces scan-line anisotropy through drift. If not accounted for, this anisotropy can dramatically impact higher spatial frequencies in certain directions, thus biasing the regression.

Challenge 2: All AFM are subject to additive Gaussian white noise (AGWN) in the form of thermal vibrations. Even after eliminating unreliable data in high frequency regions where AGWN trend overtakes signal, the differential impact of thermal noise on the retained curve will subtly under-bias the regression.

Challenge 3: Real surfaces are generally not self-affine across all scales. There generally exists a low frequency cut-off below which the PSD obeys a different trend (such as those of white or fractional Gaussian (FGN) noise). Due to the nature of the Fourier transform, spectral leakage will ensure this alternate trend subtly persists in the lower frequency portion of the self-affine region, thus over-biasing the regression.

Challenge 4: Observed RMS height varies with scan size, which results in independent pre-factors (C_0) for each scale. If not accounted for, this results in a regression converting uncertainty in the pre-factor (C_0) into uncertainty in the much more tightly-bound Hurst exponent (H).

Despite these challenges, alternative techniques for estimating the Hurst exponent provide few benefits that would justify switching to them from the log-OLS technique. Various non-PSD-based algorithms, such as box-counting or multi-return, exist for estimating the Hurst exponent from fractals, though they generally exhibit higher bias and higher variance than a PSD-based methodology [41]. Maximum Likelihood Estimation (MLE) of the Hurst exponent from the PSD, which is computationally and conceptually more expensive than the log-OLS estimation, provides marginal gains in variance, while resulting in persistent under-bias [42, 43]. Techniques which rely on applying MLE directly to real space topography images are infeasible because real surfaces are not generally fractal across all observed length scales. Wavelet-transform approaches [44, 45] can be more accurate for extremely small sample sizes if the wavelet parameters are carefully chosen, but otherwise lead to similar results and are thus beyond the scope of this work. Despite all these approaches for estimating the Hurst exponent, several open questions still remain about the accuracy of the computed values of the Hurst exponent, especially when obtained from AFM data.

Here, we quantify the impacts of the challenges outlined above on the bias in the log-OLS Hurst estimate using a Monte Carlo approach applied to synthetic AFM topography images of known input H and C_0. Synthetic AFM images were generated using a Fourier-filtering algorithm [46] with a novel scan-line anisotropy modeling technique (see Fig. S1, Supplemental Material), and simulated thermal noise through corruption post-inverse-Fourier-transform with per-pixel AGWN. Our analysis demonstrates that bias is significant and persistent, and that both the magnitude and direction of the bias varies with the scan size of the AFM image. We also explore the mitigation of this bias through stitching together multiple scan sizes into a master PSD. For this latter scenario we first address the issue of independent pre-scale C_0 due to varying observed RMS height. Furthermore, we optimize for the trade-off between having many images at few scan sizes (minimizing variance) and few images at many scan sizes (minimizing bias) for several simulated AFM experiments with a fixed number of total images collected.

II. RESULTS AND DISCUSSION

While there are many approaches to compute the PSD, in this work we relied solely on the isotropic PSD (ϕ^{iso}), which is the pseudo-1D radial average derived from a 2D map of the square modulus of the area-normalized Fourier transform of a post-processed AFM topography

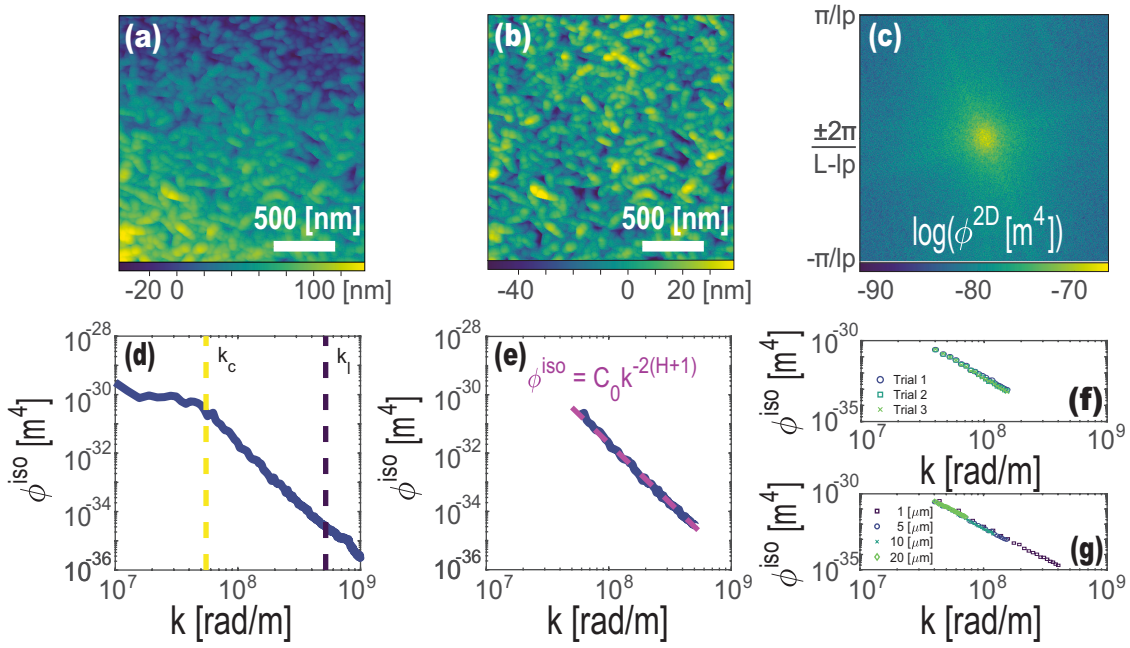

FIG. 1. Schematic diagram of procedure for estimating Hurst exponent from AFM images through power law fitting of the PSD. (a) Raw AFM image of boehmitized aluminum. (b) AFM image after row alignment and background subtraction. (c) 2D Power Spectral Density (periodogram estimate) of the AFM image. (d) Radial average of approximately isotropic 2D PSD plotted on log-log axes. Critical (k_c) and noise-threshold (k_l) wavenumbers are identified. (e) Truncated self-affine region with OLS fit to estimate the Hurst exponent. Multiple images (f) and scales (g) can be incorporated to improve accuracy of estimation.

image. Raw AFM topography images were first corrected for tilt and scan line anisotropy in real space, then windowed, and finally converted to reciprocal space using the 2D fast Fourier transform algorithm. For isotropic surfaces, the area-normalized square modulus of the 2D Fourier transform (the 2D-PSD (ϕ^{2D})) is a radially symmetric spike and thus can be averaged in rings along the radial wavenumber $\left(k = \sqrt{k_x^2 + k_y^2}\right)$. For Gaussian random surfaces the expected distribution for ϕ^{2D} is a chi-squared distribution with two degrees of freedom [47, 48], i.e., an exponential distribution, and so ϕ^{iso}, as a simple arithmetic average, is the best unbiased estimator of central tendency, and the distribution of ϕ^{iso} can be assumed to be approximately Gaussian. Binning for the radial average is an important concern as the Fourier grid is discretized linearly in reciprocal space but the final linear regression is performed on a doubly-logarithmic representation. We chose to use a linear binning scheme as it alleviates heteroscedasticity (particularly at intermediate wavenumbers) from averaging together sufficiently distinct rings. Usually several qualitative regions are visible on any given ϕ^{iso} curve, and relevant cutoffs must be identified to isolate the self-affine region present. The self-affine model can then be used to fit the data, and further data from additional images at the same or other scales can be incorporated to improve the regression. This process remains largely the same whether the source topography image is real or synthetic (see Fig. 1 for a schematic representation).

To ensure that our Monte Carlo simulations were physically meaningful, we first collected real AFM topography images of boehmitized aluminum γ-AlO(OH) [49] as well as ultra nano-crystalline diamond (UNCD). The experimental AFM topography maps could be used to identify key qualitative and quantitative features of the PSDs to be reproduced in our synthetic topography images. AFM images were collected in non-contact/tapping mode with free air amplitude tuned to avoid any degradation (i.e., wear) of the tip or surface [50]. Tip shape stability was verified using the blind tip reconstruction method [51] with UNCD. Key features in the PSDs we targeted for reproduction included the presence of two qualitatively distinct regions, i.e., one modeled as self-affine and the other as fractal Gaussian noise. The identification of these distinct regions allowed for the determination of the critical wavenumber (k_c) separating these two region, the appropriate amount of white noise at low wavelengths, and the average *image* h_{RMS} from which the model pre-factor (C_0) could be mathematically computed. Previous work has generally relied on qualitatively estimating the value of k_c [10, 11]. As this constitutes a crucial parameter in our model, we developed a novel approach to identify it with quantitative rigor using the Chow test statistic [52] (see Fig. S2, Supplemental Material). Figure 2 presents a comparison of our real and synthetic images. The good agreement between the calculated value of k_c and the average grain size along the principal axis (Fig. 2) suggests

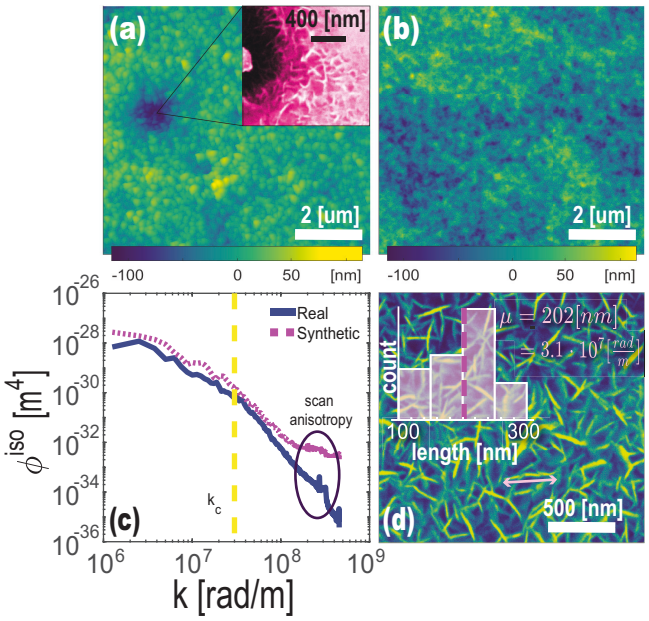

FIG. 2. (a) AFM image of boehmitized aluminum. Inset: area imaged by SEM (micrograph displayed in (d)). (b) Synthetic AFM image generated by Fourier filtering algorithm with drift anisotropy modeling and per pixel additive Gaussian white noise. (c) Comparison of the PSDs for the two images displayed in (a) and (b). The comparison demonstrates good reproduction (in synthetic images) of key features observed in experimental AFM topographic maps, particularly in the self-affine region. Slight vertical offset added to synthetic curve for clarity. Circled region highlights peak of increased power due to scan-line anisotropy. Simulated images has purposely more white noise than typical AFM images experimentally acquired in this work. (d) SEM image of boehmitized aluminum. Overlaid is the distribution of lengths of the needle-shaped grains (highlighted by double-headed arrow) as measured in ImageJ. The mean of the distribution approximately matches the value of the critical wavenumber in (c).

that the grain size is the physical limit of the "memory" of any potentially self-affine atomic processes. This observation is inline with the conclusions of recently published works [53].

Using a Monte Carlo approach, the synthetic AFM topography images prepared with a *known* input Hurst exponent (H) could be used to accurately evaluate the bias (and not just the variance) in the Hurst exponent extracted from the PSD analysis schematically outlined in Fig. 1.

A. Bias in AFM images collected at a single image scale

The collection of multiple AFM images at the same scale for a given surface can be thought of as an application of the Bartlett or Welch modified periodogram method [28], which provides an asymptotically unbiased estimator of ϕ with improved variance performance over the ordinary periodogram. Here, we are mainly interested in determining the estimator of a parameter (H) for our model of ϕ. Thus, to evaluate our estimation accuracy, we applied a Monte Carlo technique and generated a set amount of synthetic AFM images of known input H. Then, we compared the output of the log-log OLS Hurst estimation procedure (see Fig. 1) to the known input H. The results (Fig. 3) indicated a significant convergence in variance of the Hurst estimator within the first 10 images, and little further improvements beyond 100 images. Notably, the $\hat{H}$ was *not* asymptotically unbiased in our simulations, converging to values which significantly differed from the input H, and varied depending on the scale of the images. The origin of this bias can be ascribed to the presence of transition regions at the low and high wavenumber cut-offs of the observed self-affine region. At low wavenumbers, spectral leakage (i.e., redistribution of power to adjacent bins) from the fractal Gaussian noise trend steepens the regression slope and thus biases $\hat{H}$ towards smaller values (the regression slope and $\hat{H}$ have inverse directionality because $m = -2(H+1)$). At high wavenumbers, the persistent additive white noise component shallows out the regression slope and thus biases $\hat{H}$ towards larger, potentially non-physical, values. Both these effects remain present even after eliminating data beyond reliability cut-offs proposed in recent studies reported in the literature (see Jacobs et al. [10] for approaches to identifying reliability cut-offs based on instrument white noise and tip size curvature). To improve the determination of cut-off values, in the present study we used the Chow Test F statistic [52]. Fig. 3 presents both the rapid convergence of variance in $\hat{H}$, and the mechanisms responsible for bias in $\hat{H}$.

The differences in bias across scales can be attributed to changes in relative contribution of these two effects as different regions of the multi-scale PSD curve are probed. In particular, the pixel size, the pixel density, and the value of the critical wavenumber were all found to have an effect on the bias in Monte Carlo simulations. These effects can be unified onto a single curve by indexing them according to the length of the pre-self-affine region that has been probed (a metric referred to as $k_c dist$ in this work. $k_c dist$, which mathematically is the difference between $\log(k_c)$ and $\log(k_{min})$, can be negative if the pre-self-affine trend is no longer observable). However, separation of different pixel densities occurs at high $k_c dist$ values due to differing amounts of information available in the sparsely probed self-affine region. This separation can be further eliminated by normalizing the $k_c dist$ metric against a measure of noise-free self-affine region information available, which we referred to as $k_l dist$ (mathematically equal to the difference between $\log(k_l)$ and either $\log(k_c)$ or $\log(k_{min})$ - whichever is larger). The combined bias curve is presented in Fig. 4.

While the bias curve is consistent in our simulations across a range of input Hurst exponents, pixel densities, pixel sizes, and input critical wavenumbers, it

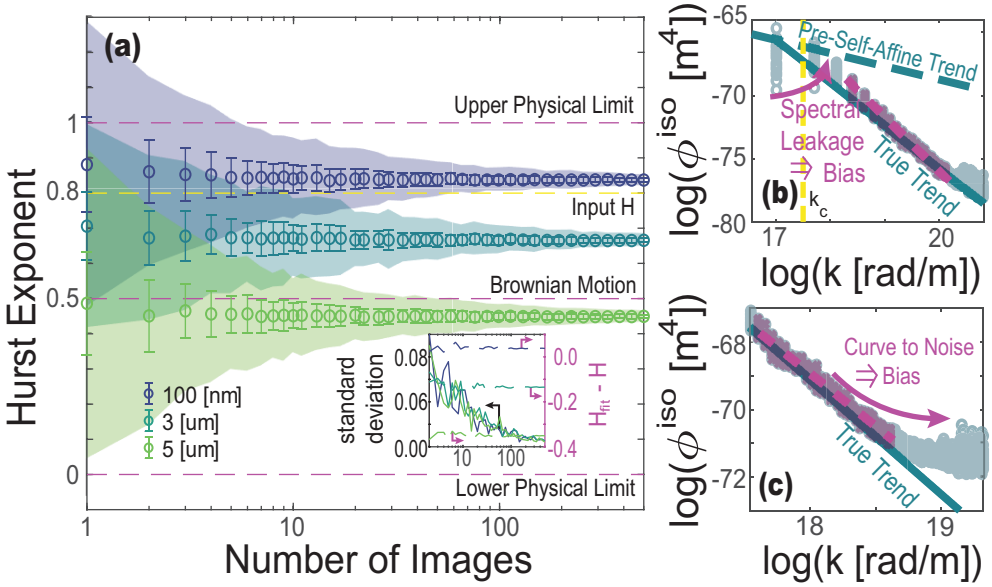

FIG. 3. (a) Repeated Monte Carlo simulations with varying numbers of total generated images. The results indicate a consistent converge in variance within 100 images, but disparate final H estimates which significantly vary depending on the image scale that was simulated. Points represent the mean of a set of 100 realizations with N images, while error bars represent the standard deviation of the set. Shaded region shows the 3-sigma confidence level. Inset shows standard deviation decreasing with N images at the same rate for all simulated image scales, but bias remaining constant and distinct. (b) At large scales, spectral leakage from the pre-self-affine trend results in steepening the fitted regression as compared to the true input trend. (c) At small scales, the persistence of subtle noise contributions results in shallowing-out the fitted regression.

strongly depends on $k_c dist/k_l dist$, especially at small $k_c dist/k_l dist$ values. These values represent what in practice is the most desirable information, i.e., images probing the smallest scales accessible to the AFM, and observing mostly or only the self-affine behavior. This finding clearly suggests that using a single image scale to estimate the Hurst exponent can be potentially misleading, while calling for the acquisition of several topography maps for achieving unbiased estimation.

B. Correctly combining AFM images at multiple scales to minimize bias and optimize bias-variance trade-off

AFM measurements allow probing multiple length scales by stitching together topographic maps. This enables for the reconstruction of a large portion of the multi-scale PSD [54–56]. For rough surfaces measured by AFM and subject to practical considerations, the largest scan size is in the order of a few 10s of µm (20 μm in the case of the instrument used in this work), while the lateral resolution is limited by the finite radius of the scanning tip ($\sim 5-10\ nm$), which corresponds to an accessible range of angular wavenumbers of $10^5 - 10^9\ \left[\frac{rad}{m}\right]$ [32]. Extending the length of the probed self-affine region can improve the accuracy of Hurst estimation. In particular, when stitching together multiple PSDs computed from topographic images acquired at different scales, achieving a good overlap of the extreme portions of the PSDs can help to significantly minimize the end-effects responsible for the bias observed in the previous section, as already pointed out by Duparre et al. [54]. Some common pitfalls addressed elsewhere in the literature, which can cause poor overlap, include improper or absent tilt-correction, improper or absent image windowing, instrument noise, and tip-size artefacts (see Jacobs et al. [10] for a comprehensive discussion of these sources of error and how to mitigate them). An additional important factor to be considered is the variation in observed h_{RMS} with image scale (Fig. 5(a)), which occurs due to sampling bias against large features when dealing with topographic maps of progressively smaller scan sizes. Since C_0 is mathematically related to h_{RMS} [10], variations in h_{RMS} can result in PSD curves acquired at different scales having different intercepts, and thus appearing as approximately parallel lines when overlaid (Fig. 5(b-c)). If differences in intercepts are not accounted for by adding degrees of freedom to the linear model, the uncertainty in the estimated intercept, C_0, can propagate to uncertainty in the estimated slope and the Hurst exponent, which has a very narrow range of physically meaningful values.

While the acquisition of multiple AFM images at different length scales can aid in improving the accuracy in the estimation of the Hurst exponent, the relative slow rate of scanning during AFM imaging [10, 42] together with the need to maximize the value of AFM instrument time require balancing the number of images being collected

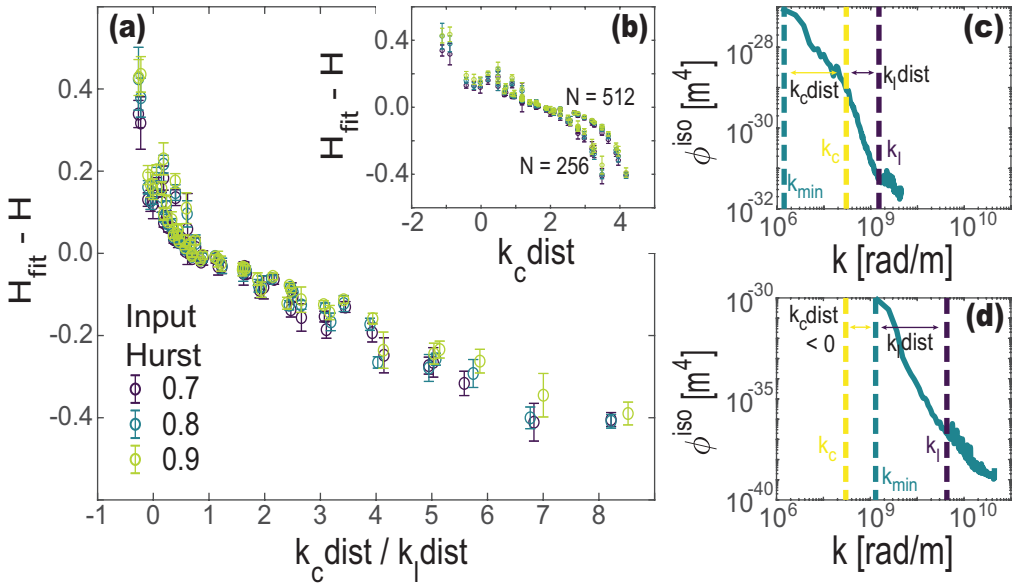

FIG. 4. (a) Absolute bias in Hurst exponent estimated from averaging 20 synthetic images as a function of the logarithmic distance between the minimum observed wavenumber and the critical wavenumber, normalized by the logarithmic distance of the noise-free self-affine region (i.e., the critical wavenumber ratio). Inset (b) shows the same curve without the normalization of the ordinate. Between trials the pixel density (256 or 512), pixel size (from $10nm$ to $1nm$), and critical wavenumber (from $8 \times 10^6 \left[\frac{rad}{m}\right]$ to $8 \times 10^7 \left[\frac{rad}{m}\right]$) were varied. Points represent the mean of a set of 10 realizations of 20 images with the same settings, while error bars represent the standard deviation of the set. The data in the inset indicates curve separation at small wave numbers (large k_c distances) for trials with varying pixel density, while on the normalized main curve this separation is eliminated. (c) Schematic of PSD with critical wavenumber ratio overlaid. (d) At small scan sizes the critical wavenumber ratio can become negative as the observed minimum wave number becomes larger than the critical wavenumber.

with the minimization of the variance in the Hurst exponent. This can be achieved by collecting many measurements in the same range of the multi-scale PSD or minimizing the bias by exploring as much of the multi-scale PSD as possible. To address this bias-variance trade-off optimization problem, we simulated multiple AFM experiments with a fixed total number of synthetic images of known input H divided across multiple scales in different ways. We then compared the bias and variance (in the form of the mean square error criterion) across ten repeats of each synthetic experiment. Two approaches were used to progressively incorporate many scales. For the constant spacing (CS) approach, we began with the midpoint of the scales practically available to the AFM (arbitrarily chosen to be 5 [μm] scan size, 256x256 pixel image), and used additional scales to probe further out from the middle-scale while maintaining constant spacing between scales. Thus, for 2 scales we used 8 [μm] and 5 [μm] scan sizes, for 3 scales we used 8, 5 and 3 [μm] scan sizes, and for 4 scales we used 10, 8, 5 and 3 [um] scan sizes. For the constant range (CR) approach, we *began* with the entire range of scales practically available to the AFM (from a 20 [μm] scan size to a 0.2 [μm] scan size, 256x256 pixel image), and used additional scales to fill-in the range. Both approaches became equivalent at the limiting conditions of 1 scale, and the max number of scales the simulation was designed for, i.e. 12 scales. We also evaluated different number of total images (i.e., high (120), medium (24) and low (12)). The number of total images was selected based on the variance convergence analysis from the previous section, while maintaining even divisibility with as many potential N scales as possible. Input C_0 for each simulated image was varied by following the relationship between h_{RMS} and scan size observed in our real samples, while including standard Gaussian noise of variance 0.1. Estimated Hurst output was determined using a procedure that accounted for the differences in input C_0 by assuming an independent intercept (and shared slope) for each scale. The results of these bias-variance trade-off experiments are presented in Fig. 5.

The simulation results suggest that, for minimizing bias in the estimation of the Hurst exponent, probing the entire range of scales practically accessible to the AFM is superior to probing a narrow but closely spaced range. Moreover, beyond an optimum of 3-6 scales (strict optimum at 4), it is a more productive use of limited AFM time to collect more images at each of the probed scales, than it is to probe additional scales. Mean square errors below 3% or ± 0.024 for a Hurst exponent of 0.8 can feasibly be achieved with as few as 12 total images across 4 scales.

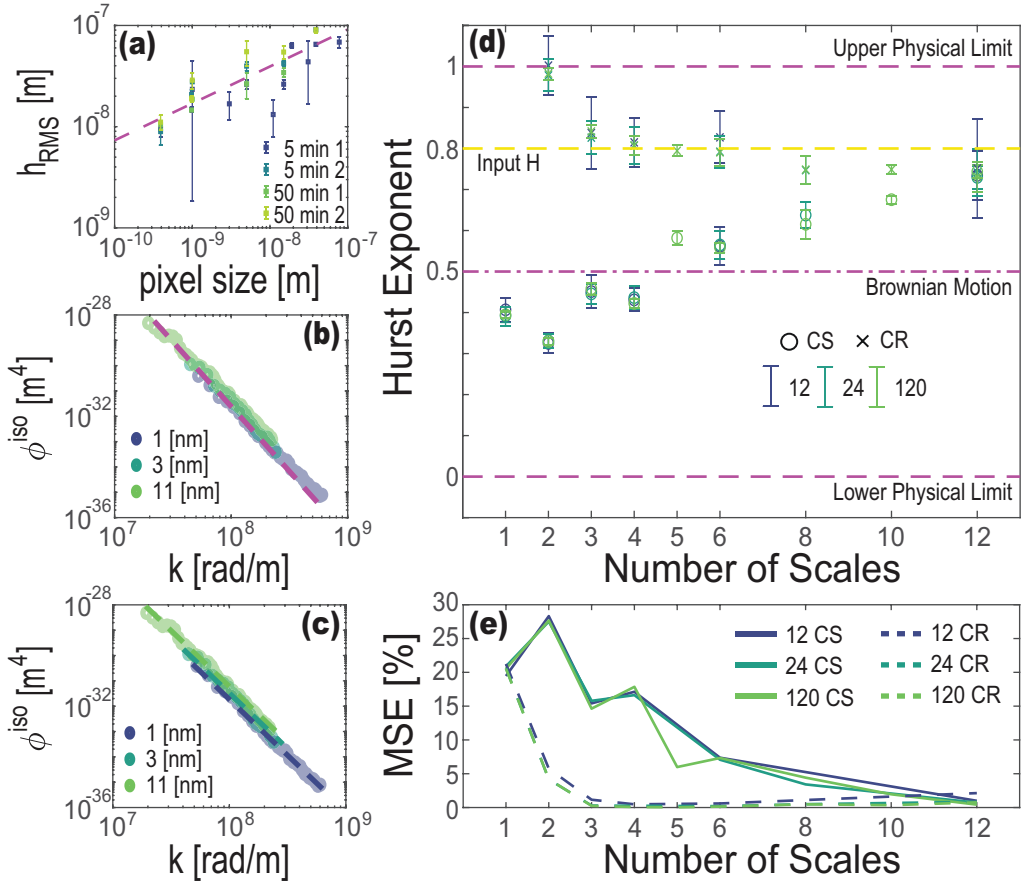

FIG. 5. (a) Dependence of h_{RMS} on image scale across four boehmitized aluminum samples prepared by immersing aluminum in boiling water for varying lengths of time. Overlaid trend was used for varying input C_0 with scale in the simulation. (b) Schematic of "stair-step" behavior that can occur due to slight variations in the observed RMS height across different scales, with a poor treatment (fitting to all points combined) overlaid. Data is from an aluminum sample immersed in boiling water for 5 min. (c) Schematic of correct treatment of "stair-step" behavior, with independent intercepts for each scale. (d) Monte Carlo simulation of bias-variance trade-off that occurs when incorporating more scales with fixed total number of collected images. Points represent the mean of a set of 10 realizations with either 12, 24 or 120 total images divided evenly across N scales, while error bars represent the standard deviation of the set. Circular points denote the constant spacing (CS) approach to adding more scales, while cross points denote the constant range (CR) approach. For 4 and 10 scales, only realizations with 120 total images were generated, while for 8 scales, only realizations with 24 or 120 total images were generated. (e) MSE of bias-variance trade-off in (e) showing optimum between 3-6 scales using the constant range (CR) approach.

III. CONCLUSIONS

In the present work, we generated synthetic AFM images of known input Hurst exponent through a Fourier filtering algorithm and evaluated, using a Monte Carlo approach, the presence of bias in the Hurst exponent estimated on the basis of ordinary-least-squares (OLS) regression of a doubly logarithmic transformation of the power spectral density (PSD). The PSD of the synthetic images was designed to reproduce typical PSD curves computed from real, measured boehmite and UNCD surfaces. We evaluated the bias under two different scenarios, namely many images collected at the same scale, and multiple overlapping scales stitched together to create a master PSD.

For the single-scale scenario we identified that while variance converges rapidly - within $\sim$ 10 images with little further improvement beyond 100 images - bias is significant and persistent, and both the magnitude and direction of the bias varies with the scan size of the AFM image. We also identified the mechanisms responsible for this bias and used that knowledge to collapse the bias results under varying conditions onto a single predictive bias curve. This single predictive bias curve indicated a significant increase in bias at the most desirable scales, i.e., those at the limit of what is accessible to the AFM. This finding highlights the potential limitation of using only a single image scale to estimate the Hurst exponent.

For the multiple scale scenario, we optimized for the variance-bias trade-off . The results indicated that wide ranges in image scales are desirable for minimizing bias, and that mean square errors below 3% can feasibly be

achieved with as few as 12 total images across an optimum of 4 scales.

Finally we observed evidence in our measurements of real surfaces supporting a link between the critical wavenumber and the grain size, and we proposed a novel technique for rigorously and quantitatively determining the value of the critical wavenumber by maximizing the Chow test statistic.

IV. MATERIALS AND METHODS

A. Sample Preparation

Aluminum 6061 (McMaster-Carr, Douglasville, GA, USA) discs were mechanically ground and polished on a LaboPol-20 (Struers USA, Cleveland, OH, USA) down to a final $1/4\mu m$ diamond paste polish. Once prepared, the samples were sonicated in acetone and then submerged in boiling ultrapure water (resistance 18.2 MΩ) to form a nano-rough boehmite (γ-AlO(OH)) film following the procedure outlined in Ref. [49]. The size and density of the boehmite crystals were manipulated by varying the duration of the boehmitization process from 5 minutes to over 2 hours.

B. AFM Experiments

Topography images were obtained with an XE100 AFM (Park Systems, Suwon, South Korea) in true non-contact (TM) mode using PPP-NCHR probes (Nanosensors, Neuchatel, Switzerland). Statistical roughness information such as h_{RMS} values (Figure S3) were derived from topography images obtained with an MFP-3D Origin+ AFM (Oxford Instruments Asylum Research, Santa Barbara, CA, USA) in tapping mode using HQ:NSC15 probes (MikroMasch USA, Watsonville, CA, USA). Spring constants were calibrated using Sader's method [57]. The tapping mode parameters (such as the free oscillation amplitude) were tuned so that the peak contact stress would never exceed the yield stress of either the tip or the substrate [50]. Absence of tip degradation was verified by performing blind-tip reconstruction [51] against an ultra-nanocrystalline diamond (UNCD, Aqua 25, Advanced Diamond Technologies, USA) test substrate both before and after each experiment. AFM scanning rates varied from 1 Hz for smaller images ($\tilde{}1\mu m^2$) to 0.1 Hz for larger images ($\tilde{}20\mu m^2$). All substrates were sonicated in acetone before acquiring topography images. AFM image post-processing was performed in Gwyddion using the polynomial background subtraction [58] subroutine to account for tilt and the align rows [59] subroutine to minimize the impact of scanline anisotropy.

C. Synthetic Image Generation

Synthetic self-affine images were generated using an inverse-Fourier filtering algorithm [46]. Fourier transform coefficients were generated with a uniformly distributed random phase and a normally distributed absolute square with mean and standard deviation following either a fractal Brownian noise trend [26, 27] for wavevectors above k_c or a fractal Gaussian noise trend [60, 61] for wavevectors below k_c. The pre-exponential constant C_0 was computed [10] based on a desired output h_{RMS} matching experimental results in the range 20-40nm. To simulate drift related to the scanning speed anisotropy in AFM (originating from the presence of a fast- and a slow-scan direction), the Fourier coefficients were additionally scaled following an empirical half-Gaussian trend, to replicate the anisotropy visible on real data in Fig. 2. Accurate reproduction required scaling only those directions for which the expected amount of AFM drift would be similar to total distance traveled by the AFM tip (see Fig. S1 Supplemental Material). To mimic thermal noise, the resultant image after inverting the FFT was corrupted with per-pixel additive Gaussian white noise (zero mean, standard deviation equal to the expected magnitude of the self-affine power law at twice the pixel length). Images generated by this algorithm are by nature periodic.

D. Spectral Analysis

Images were first windowed using the radially symmetric Hann window, then normalized using squared-window normalization to preserve RMS-roughness sum rules [10]. Windowed images were Fourier transformed using MATLAB's built-in 2D-FFT routine [62]. The resultant FT value was squared to obtain the periodogram estimate of the power spectral density [28]. Coordinates in reciprocal space were computed treating the position of each pixel as being the center of the pixel. The radial average of the 2D Fourier grids was computed while remaining in the linear binning scheme. For the resultant isotropic PSDs, the critical wavenumber and noise-threshold cut-offs were determined by maximizing the Chow test statistic [52]. Initial estimates for these cut-offs were obtained through qualitative appraisal, or through a combined white noise threshold [10], and anisotropy based approach, respectively.

ACKNOWLEDGMENTS

We thank Gustavo de Veciana and Alan Bovik for insightful discussions about the statistics of colored noises. We thank Petre Stoica for graciously sending us a copy of the solutions manual to his book *Spectral Analysis of Signals*. The material is based upon work supported by the Welch Foundation (Grant No. F-2002-20190330), the National Science Foundation Faculty Early Career

Development Program (Grant No. 2042304), and the Taiho Kogyo Tribology Research Foundation (Grant No. 20A03). F.M. acknowledges support from the 2018 Ralph E. Powe Junior Faculty Enhancement Award sponsored by the Oak Ridge Associated Universities (ORAU), and from the Walker Department of Mechanical Engineering and the Texas Materials Institute at the University of Texas at Austin. R.C acknowledges support from the National Science Foundation through the 2020 Graduate Research Fellowship Program (Fellow ID. 2020308624), and from the College of Engineering at the University of Texas at Austin.

[1] F. W. DelRio, M. P. de Boer, J. A. Knapp, D. E. J. Reedy, P. J. Clews, and M. L. Dunn, The role of van der waals forces in adhesion of micromachined surfaces., Nature Materials **4**, 629 (2005).

[2] L. Pastewka and M. O. Robbins, Contact between rough surfaces and a criterion for macroscopic adhesion, PNAS **111**, 3298 (2014).

[3] M. H. Müser, A dimensionless measure for adhesion and effects of the range of adhesion in contacts of nominally flat surfaces, Tribology International **100**, 41 (2016).

[4] P. Lonardo and G. Micheletti, Measurement of smooth surface roughness by means of photometric methods., CIRP Annals **23**, 189 (1974).

[5] R. L. Webb., Kern lecture award paper: Odyssey of the enhanced boiling surface., Journal of Heat Transfer **126**, 1051 (2004).

[6] C. Sodtke and P. Stephan, Spray cooling on micro structured surfaces., International Journal of Heat and Mass Transfer **50**, 4089 (2007).

[7] G. Pugliese, S. Tavares, E. Ciulli, and L. Ferreira, Rough contacts between actual engineering surfaces: Part ii. contact mechanics, Wear **264**, 1116 (2008).

[8] A. Bruzzone, H. Costa, P. Lonardo, and D. Lucca, Advances in engineered surfaces for functional performance, CIRP Annals **57**, 750 (2008).

[9] A. Vakis, V. Yastrebov, J. Scheibert, L. Nicola, D. Dini, C. Minfray, A. Almqvist, M. Paggi, S. Lee, G. Limbert, J. Molinari, G. Anciaux, R. Aghababaei, S. Echeverri Restrepo, A. Papangelo, A. Cammarata, P. Nicolini, C. Putignano, G. Carbone, S. Stupkiewicz, J. Lengiewicz, G. Costagliola, F. Bosia, R. Guarino, N. Pugno, M. Müser, and M. Ciavarella, Modeling and simulation in tribology across scales: An overview, Tribology International **125**, 169 (2018).

[10] T. D. Jacobs, T. Junge, and L. Pastewka, Quantitative characterization of surface topography using spectral analysis, Surface Topography Metrology and Properties **5**, 013001 (2017).

[11] A. Gujrati, S. R. Khanal, L. Pastewka, and T. D. Jacobs, Combining tem, afm and profilometry for quantitative topography characterization across all scales, ACS Applied Materials Interfaces **10**, 29169 (2018).

[12] S. Akarapu, T. Sharp, and M. O. Robbins, Stiffness of contacts between rough surfaces, Physical Review Letters **106**, 204301 (2011).

[13] C. Campañá, B. N. J. Persson, and M. H. Müser, Transverse and normal interfacial stiffness of solids with randomly rough surfaces, Journal of Physics: Condensed Matter **23**, 085001 (2011).

[14] R. Pohrt and V. L. Popov, Normal contact stiffness of elastic solids with fractal rough surfaces, Physical Review Letters **108**, 104301 (2012).

[15] L. Pastewka, B. L. Prodanov, M. H. Müser, M. O. Robbins, and B. N. J. Persson, Finite-size scaling in the interfacial stiffness of rough elastic contacts, Physical Review E **87**, 062809 (2013).

[16] R. N. Wenzel, Surface roughness and contact angle., Journal of Physical Chemistry **53**, 1466 (1949).

[17] A. B. D. Cassie and S. Baxter, Wettability of porous surfaces, Transactions of the Faraday Society **40**, 546 (1944).

[18] D. Quéré and M. Reyssat, Non-adhesive lotus and other hydrophobic materials, Philosophical Transactions of the Royal Society A **366**, 1539 (2008).

[19] T.-S. Wong, S.-H. Kang, K. Y. Tang, E. J. Smythe, B. D. Hatton, A. Grinthal, and J. Aizenberg, Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity., Nature **477**, 443 (2011).

[20] B. Gotsmann and M. A. Lantz, Quantized thermal transport across contacts of rough surfaces, Nature Materials **12**, 59 (2013).

[21] B. N. J. Persson, Theory of rubber friction and contact mechanics, Journal of Chemical Physics **115**, 3840 (2001).

[22] B. N. J. Persson, Elastoplastic contact between randomly rough surfaces, Physical Review Letters **87**, 116101 (2001).

[23] G. O. Berim and E. Ruckenstein, Nanodrop on a nanorough hydrophilic solid surface: Contact angle dependence on the size, arrangement, and composition of the pillars, Journal of Colloid and Interface Science **359**, 304 (2011).

[24] R. Enright, N. Miljkovic, J. L. Alvarado, K. Kim, and J. W. Rose, Dropwise condensation on micro- and nanostructured surfaces, Nanoscale and Microscale Thermophysical Engineering **18**, 223 (2014).

[25] E. Bormashenko, Progress in understanding wetting transitions on rough surfaces, Advances in Colloid and Interface Science **222**, 92 (2015).

[26] B. Mandelbrot, How long is the coast of britain? statistical self-similarity and fractional dimension, Science **156**, 636 (1967).

[27] B. Mandelbrot, Self-affine fractals and fractal dimension, Physica Scripta **32**, 257 (1985).

[28] P. Stoica and R. Moses, *Spectral Analysis of Signals* (Prentice Hall Inc., 2004).

[29] K. El-Shennawy, *Communication Theory and Signal Processing for Transform Coding* (Bentham Science Publishers, 2018).

[30] S. Solhjoo and A. I. Vakis, Surface roughness of gold substrates at the nanoscale: An atomistic simulation study, Tribology International **115**, 165 (2017).

[31] C. Gerber and H. P. Lang, How the doors to the nanoworld were opened, Nature Nanotechnology **1**, 3 (2006).

[32] E. Meyer, H. J. Hug, and R. Bennewitz, *Scanning Probe Microscopy: The Lab on a Tip* (Springer, 2004).

[33] P. J. de Groot, Interference microscopy for surface structure analysis, in *Handbook of Optical Metrology*, edited by T. Yoshizawa (CRC Press, 2017) Chap. 31, pp. 791–824.

[34] P. J. de Groot, Principles of interference microscopy for the measurement of surface topography, Advances in Optics and Photonics **7**, 1 (2015).

[35] R. S. Sayles and T. R. Thomas, Surface topography as a nonstationary random process, Nature **271**, 431 (1978).

[36] M. Hasegawa, J. Liu, K. Okuda, and M. Nunobiki, Calculation of the fractal dimensions of machined surface profiles, Wear **192**, 40 (1996).

[37] T. Gneiting and M. Schlather, Stochastic models that separate fractal dimension and the hurst effect, SIAM Review **46**, 269 (2004).

[38] H. Jummomem, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, Methods for imputation of missing values in air quality data sets, Atmospheric Environment **38**, 2895 (2004).

[39] A. R. Rao and D. Bhattacharya, Comparison of hurst exponent estimates in hydrometeorological time series, Journal of Hydrologic Engineering **4**, 10.1061/(ASCE)1084-0699(1999)4:3(225) (1999).

[40] M. Tarnopolski, N. Żywucka, V. Marchenko, and J. Pascual-Granado, A comprehensive power spectral density analysis of astronomical time series. i. the fermilat gamma-ray light curves of selected blazars, The Astrophysical Journal **250**, 10.3847/1538-4365/aba2c7 (2020).

[41] J. Schmittbuhl, J.-P. Vilotte, and S. Roux, Reliability of self-affine measurements, Physical Review E **51**, 131 (1995).

[42] C.-Y. Wen and R. Acharya, Self-similar texture characterization using a fourier-domain maximum likelihood estimation method, Pattern Recognition Letters **19**, 735 (1998).

[43] C. Kendziorski, J. Bassingthwaighte, and P. Tonellato, Evaluating maximum likelihood estimation methods to determine the hurst coefficient, Physica A: Statistical Mechanics and its Applications **273**, 439 (1999).

[44] M. Fadili and E. Bullmore, Wavelet-generalized least squares: A new blu estimator of linear regression models with 1/f errors, NeuroImage **15**, 217 (2002).

[45] I. Simonsen, A. Hansen, and O. M. Nes, Determination of the hurst exponent by use of wavelet transforms, Physical Review E **58**, 2779 (1998).

[46] D. Saupe, Algorithms for random fractals, in *The Science of Fractal Images*, edited by H.-O. Peitgen and D. Saupe (Springer, 1988) Chap. 2, pp. 71–136.

[47] J. Huillery, F. Millioz, and N. Martin, On the description of spectrogram probabilities with a chi-squared law, IEEE Transactions on Signal Processing **56**, 2249 (2008).

[48] R. A. Kulak, Accuracy and repeatability of noise measurements with a discrete fourier transform, in *2008 IEEE Dallas Circuits and Systems Workshop: System-on-Chip - Design, Applications, Integration, and Software* (2008) pp. 1–4.

[49] P. Kim, M. J. Kreder, J. Alvarenga, and J. Aizenberg, Hierarchical or not? effect of the length scale and hierarchy of the surface roughness on omniphobicity of lubricant-infused substrates, Nano Letters **13**, 1793 (2013).

[50] V. Vahdat and R. W. Carpick, Practical method to limit tip–sample contact stress and prevent wear in amplitude modulation atomic force microscopy, ACS Nano **7**, 9836 (2013).

[51] J. S. Villarrubia, Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation, Journal of Research of the National Institute of Standards and Technology **102**, 425 (1997).

[52] G. C. Chow, Tests of equality between sets of coefficients in two linear regressions, Econometrica **28**, 591 (1960).

[53] A. Gujrati, A. Sanner, S. R. Khanal, N. Moldovan, H. Zeng, L. Pastewka, and T. D. B. Jacobs, Comprehensive topography characterization of polycrystalline diamond coating, Surface Topography: Metrology and Properties **9**, 014003 (2021).

[54] A. Duparré, J. Ferre-Borrull, S. Gliech, G. Notni, J. Steinert, and J. M. Bennett, Surface characterization techniques for determining the root-mean-square roughness and power spectral densities of optical components, Applied Optics **41**, 154 (2002).

[55] T. Candela, F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, and E. E. Brodsky, Roughness of fault surfaces over nine decades of length scales, Journal of Geophysical Research Solid Earth **117**, 1 (2012).

[56] E. Marx, I. J. Malik, Y. E. Strausser, T. Bristow, N. Poduje, and J. C. Stover, Power spectral densities: A multiple technique study of different si wafer surfaces, Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena **20**, 10.1116/1.1428267 (2002).

[57] J. E. Sader, J. W. M. Chon, and P. Mulvaney, Calibration of rectangular atomic force microscope cantilevers, Review of Scientific Instruments **70**, 3967 (1999).

[58] P. Klapetek, D. Nečas, and C. Anderson, Gwyddion user guide, leveling and background, `http://gwyddion.net/documentation/user-guide-en/leveling-and-background.html` (2021).

[59] P. Klapetek, D. Nečas, and C. Anderson, Gwyddion user guide, scan line artefacts, `http://gwyddion.net/documentation/user-guide-en/scan-line-defects.html` (2021).

[60] B. Mandelbrot and J. W. van Ness, Fractional brownian motions, fractional noises and applications, SIAM Review **10**, 422 (1968).

[61] J. Beran, Y. Feng, S. Ghosh, and R. Kulik, *Long-Memory Processes, Probabilistic Properties and Statistical Methods* (Springer, 2013).

[62] Matlab documentation, 2d fast fourier transform., `https://www.mathworks.com/help/matlab/ref/fft2.html`. (2021).

[63] See supplemental material at, `http://link.aps.org/supplemental/givenbypublisher`, for an in-depth discussion of the mathematics underlying the PSD; the fractal modeling of the PSD; the Fourier-Filtering Algorithm used to generate synthetic images; the novel anisotropy modeling technique; and the Chow test optimization used to locate critical wavenumbers.

[64] O. Tricinci, T. Terencio, B. Mazzolai, N. M. Pugno, F. Greco, and V. Mattoli, 3d micropatterned surface inspired by salvinia molesta via direct laser lithography, ACS Applied Material Interfaces **7**, 25560 (2015).

[65] H.-J. Choi, S. Choo, J.-H. Shin, K.-I. Kim, and H. Lee, Fabrication of superhydrophobic and oleophobic surfaces with overhang structure by reverse nanoimprint lithography, Journal of Physical Chemistry C **117**, 24354 (2013).

[66] M. Nosonovsky and B. Bhushan, Why re-entrant surface topography is needed for robust oleophobicity, Philosophical Transactions of the Royal Society A **374**,

10.1098/rsta.2016.0185 (2016).
[67] J. W. Cooley and J. W. Tukey, An algorithm for the machine calculation of complex fourier series, Mathematics of Computation **19**, 297 (1965).
[68] J. M. Elson and J. M. Bennett, Calculation of the power spectral density from surface profile data, Applied Optics **34**, 201 (1995).
[69] K. M. M. Prabhu, *Window Functions and Their Applications in Signal Processing* (CRC PRess, 2014).
[70] B. Mandelbrot, D. E. Passoja, and A. J. Paullay, Fractal character of fracture surfaces of metals, Nature **308**, 721 (1984).
[71] H. Rott, The analysis of backscattering properties from sar data of mountain regions, IEEE Journal of Oceanic Engineering **9**, 347 (1984).
[72] M. Keshner, 1/f noise, Proceedings of the IEEE **70**, 212 (1982).
[73] B. N. J. Persson, Contact mechanics for randomly rough surfaces, Surface Science Reports **61**, 201 (2006).