Establishing nanoscale heterogeneity with nanoscale force measurements

Yun-Hsiang Chang,†,§ Tuza Olukan,†,§ Chia-Yun Lai,† Sergio Santos,† Tze-Yu Lin,† Harry Apostoleris,†

Josep Font,‡ Victor Barcons,‡ and Matteo Chiesa*,†

†Laboratory for Energy and NanoScience (LENS), Institute Center for Future Energy (iFES), Masdar Institute of Science and Technology, Abu Dhabi, UAE

‡Departament de Disseny i Programació de Sistemes Electrònics, UPC-Universitat Politècnica de Catalunya, Av. Bases, 61, 08242 Manresa, Spain

§Equally contributing authors

ABSTRACT

Establishing the presence or absence of nanoscale compositional heterogeneity with nanoscale resolution is becoming instrumental for the development of many fields of science. Force versus distance measurements and parameters directly or indirectly derived from these profiles can be potentially employed for this purpose with sophisticated instruments such as the atomic force microscope (AFM). On the other hand, standards are necessary to reproducibly and conclusively support hypothesis from experimental data and these standards are still emerging. Here, we define a set of standards for providing data originating from atomic force measurements to be employed to compare between sample properties, parameters, or, more generally, compositional heterogeneity. We show that reporting the mean and standard deviation only might lead to inconsistent conclusions. The fundamental principle behind our investigation deals with the very definition of reproducibility and repeatability in terms of accuracy and precision, and we establish general criteria to ensure that these hold without the need of restricting assumptions.

1. INTRODUCTION

A major motivation behind mapping material properties with nanoscale resolution relates to understanding the relationships between dimensions and properties, since these might allow fine-tuning of these properties. In fact, a unifying goal in nanotechnology is to exploit the properties that structures exhibit when reaching nanoscale dimensions,[1–3] and such properties and phenomena are many times interpreted in terms of nanoscale heterogeneity.[4–6] Nanoscale heterogeneity might manifest as a range of exotic mechanical,[5,7] optical,[8] chemical,[9] or other phenomena,[1] and properties might be exploited in terms of functionality[10] ranging from mechanics[7] to biology.[6,9] AFM has the capacity to map nanoscale properties with high resolution and is therefore very commonly employed to map nanoscale heterogeneity.[4,11–14] In one of its fundamental modes of operation, AFM is employed to measure the distance dependent forces that the tip of the AFM, i.e., a nanostructure, experiences when it is brought into close proximity with a surface. From this tip−surface force profile or F_{ts}, several material dependent features can be recorded. Arguably, in experiments dealing with other than mechanical contact, the most commonly employed-model independent-feature derived from the force profile is the minima in force, or force of adhesion (F_{AD}). Another useful feature that has the advantage of also being model independent is the work of adhesion or area under the curve (W_{AD}). These[6,14,15] and other features, i.e., energy dissipation,[4,16,17] stiffness,[11,18] etc., are commonly employed to establish the presence or absence of nanoscale heterogeneity in a given surface.[19] Here, we explore the experimental conditions and requirements to establish the presence or absence of nanoscale compositional heterogeneity from experimental force measurements in AFM by considering experimental errors in the context of accuracy and precision as a function of the samples' size N. We employ a homogeneous material, i.e., a sapphire surface, to exemplify and verify our hypothesis. We focus on a fundamental parameter of force profiles, namely, F_{AD}, but we exemplify how our methodology can be generalized to any other parameter extracted from force measurements by employing the same concepts to W_{AD}. Our data shows that it is possible to improve precision, i.e., decrease the interval or margin of error, while maintaining accuracy, i.e., repeatedly including in the given interval or within a margin of error the true mean of the parameter being measured, by sufficiently increasing N. We show however that this is not achievable by directly assuming that the underlying distribution is normally distributed. In particular,

applying concepts from standard theory of inference with the assumptions of normal distributions (or more generally Student's t-distribution) leads to very large errors and finally to erroneous or inconsistent conclusions. For example, even when providing mean valueswith intervals or margins of error (standard errors computed assuming normal or Student's t-distributions) obtained from 300 data points from a given location of a given homogeneous material, the intervals lie outside the range of intervals produced by subsequently and continuously acquiring 300 data points from the same material and without readjusting the experimental setup. In practice, such results would imply that a sample's surface at a given location is different from itself. This conclusion makes comparisons with other locations, phases, or samples' surfaces questionable at best and, at worst, irrelevant and calls for a more robust test of comparison between data sets. Our experiments show that averaging over at least 200−300 points might be required in order to obtain a normally distributed distribution as required by the central limit theorem.[20] In order to obtain sample sizes in the order of 10−30 data points, this would imply acquiring at least 2000−9000 data points per experiment. Since force measurements require at least fractions of a second,[11,13,14,21,22] the associated time-cost would be considerable for standard experimentation.

More importantly, we find evidence that shows inconsistencies, such as that detailed above, even when the data passes standard normality tests. In this work, we define and employ a set of metrics to deal with accuracy and precision in the context of force measurements and provide a protocol for measurements that allows us to define, with a chosen confidence, that a given measurement with N data points comes with an error, or a given precision in the measurement that makes measurements self-consistent. More thoroughly, we define a set of standards to compare between sample properties, parameters, or, more generally, compositional heterogeneity, with nanoscale force measurements or parameters derived from these. The fundamental principle behind our investigation deals with the very definition of reproducibility and repeatability in terms of accuracy and precision, and we establish general criteria to ensure that these hold without the need of restricting assumptions. We further employ a model system consisting of a copolymer to exemplify that less than 100 data points are enough to establish the presence of nanoscale heterogeneity with the concepts and metrics that we employ here and report a resolution or precision of $\sim$10 pN for F_{AD}. The methodology is also employed to exploit W_{AD} as a parameter to establish heterogeneity in order to show that the theory can be generalized to other parameters. In

the case of W_{AD}, ~200 data points are necessary to arrive to the same conclusions, i.e., to establish nanoscale heterogeneity.

2. EXPERIMENTAL METHODS

The tip−sample force F_{ts} or force profile has been reconstructed from experimental amplitude and phase distance APD curves recorded with an AFM operated in the amplitude modulation (AM) mode. A standard Cypher AFM from Asylum Research was employed to perform the experiments and cantilevers from Olympus (AC160TS) where employed throughout. The force can be reconstructed from the observables amplitude A and phase lag Φ by writing the tip−sample distance d as $d = z_c + z$, where z_c is the cantilever−surface distance and z is the cantilever deflection. The inversion from observables to force exploits the Sader−Jarvis−Katan formalism[23,24] by writing the minimum distance of approach per cycle d_m as $d_m \approx z_c − A$.

$$F_{ts}(d_m) = 2k \int_{u=d_m}^{u=\infty} \left[\left(1 + \frac{A^{1/2}(u)}{8\sqrt{\pi(u-d_m)}} \right) \Omega(u) - \frac{A^{3/2}(u)}{\sqrt{2(u-d_m)}} \frac{d\Omega(u)}{du} \right] du$$

(1)

and

$$\Omega(d_m) = \left[1 + \frac{A_0}{QA} \cos(\Phi(d_m)) \right]^{1/2} - 1$$

(2)

where k is the cantilever stiffness, Q is the Q factor, $\Omega(d_m)$ is the normalized frequency shift, and A_0 is the free (unperturbed) amplitude. The algorithm in eqs 1 and 2 has been implemented in Matlab with standard functions, and the resulting raw force F_{ts} has been smoothened with the standard rloess method (moving average filter of 30) from Matlab prior to calculating F_{AD} and W_{AD} (numerically integrated). The speed of acquisition was limited by the AFM, i.e., one amplitude and phase curve ~1 s, while 10^4 amplitude and phase curves could be processed to produce a 10^4 force distance profile in approximately 10 min.

Sapphire has been employed as a model system because it presents a homogeneous surface where F_{AD} and W_{AD} should be homogeneous even in the nanoscale. This model system has been employed to establish accuracy and precision in the measurements and the convergence to a value as the sample's size increases. Standard cantilevers from Olympus (AC160TS) with $k \approx 40$ N/m, $Q \approx 500$, and $f_0 = 300$ kHz have been employed throughout. Two parameters that can lead to systematic errors have been carefully considered. First, during a given experiment, the tip radius R might vary due to wear-typically increasing R.[25,26] Since both F_{AD} and W_{AD} increase with R,[27] wear would lead to divergence and lack of reproducibility with increasing N. In order to avoid such errors in our experiments, the tip's effective radius was monitored in situ with the critical amplitude A_c method,[25,28,29] i.e., by monitoring the minimum value of free amplitude A_0 for which attractive to repulsive transitions occur. In all experiments here, the A_c value remained constant to fractions of nm which lies within the error of the A_c method to characterize the tip radius,[28] i.e., 1−2 nm. Tip wear was physically prevented in the experiments by fine-tuning peak forces when acquiring data. In practice, peak forces where controlled and limited to 2−3 nN (data not shown) by maintaining the minimum amplitude of oscillation sufficiently high in the amplitude distance curves (APDs) as done elsewhere.[30] Second, variations in the position onto which the laser beam is reflected from the cantilever's surface might lead to slight variations in the volt to meter conversion with obvious implications for the values of force calculated from eqs 16 and 17. In order to avoid this second source of systematic error, the laser was aligned and adjusted for at least 30 min prior to acquiring data. This was done to avoid maladjustments during the experiments due to thermal expansion of the microcantilever or other. Then, the laser beam was never readjusted during the experiments.

The F_{AD} between a sphere, i.e., the tip, and a flat surface is typically accepted to be proportional to the sphere's radius R[27,31]
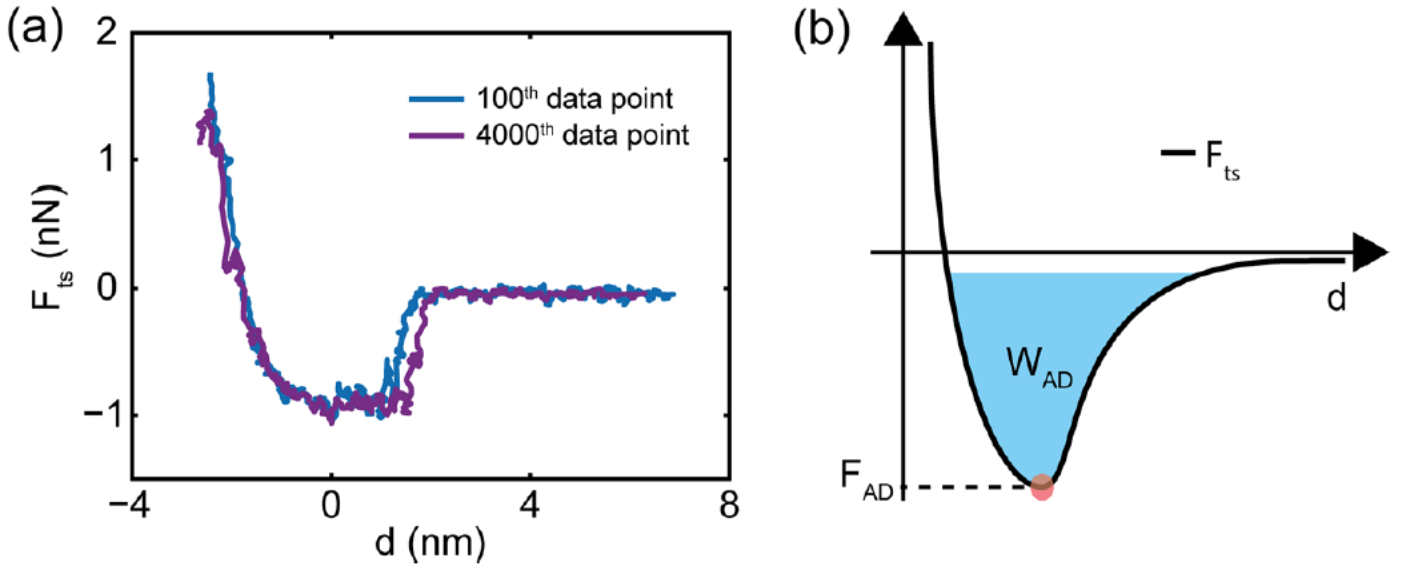
$$F_{AD} = 4\pi R \gamma \tag{3}$$

Figure 1. (a) Two experimental force profiles acquired on a sapphire surface. Blue line: n = 100. Purple line: n = 4000. (b) Illustration of the method to compute F_{AD} and W_{AD} (area under the curve).

where γ is the surface tension.[27] In our work, F_{AD} is calculated directly from experimental force profiles as illustrated in Figure 1 (continuous lines) and it is identified as minima in force (Figure 1b). Thus, as opposed to the expression in eq 3, the parameter F_{AD} as defined in this work is model independent. Considering eq 3 however is still constructive in order to get physical intuition for the meaning of minima in force. W_{AD} is also directly computed from the experimental force profile as the area under the force curve

$$W_{AD} = \int_{d_0}^{\infty} F_{ts}dx \qquad (4)$$

where the integral goes from $x = d_0$, i.e., $F_{ts} = 0$ at d_0 on the left side of the forces in Figure 1, to infinity. Again, for practical reasons and for easy computation, we have defined W_{AD} as the area under the line $0.05F_{AD}$ (Figure 1b). Profiles similar to those in Figure 1a were collected on a sapphire surface at a constant rate of 0.5 Hz, i.e., one force profile for every 2 s, and at a given location on the surface by following the procedures described above. Further details on the methodology for force reconstruction in AM AFM are provided elsewhere.[24,32]

Sets of data of 5000 points or more were acquired continuously and for hours using such AM AFM force reconstruction methodology.

As an example, two force profiles have been plotted in Figure 1a. One of them is the 100th data point, or n = 100 (blue line), and the other one belongs to the same set of data (5000 points) for n = 4000 (purple line); i.e., data points are separated in time by at least 2 h. The fact that $F_{AD} \approx -1$ nN for n = 100 and n = 4000 (more than 2 h of collecting data) provides some intuitive evidence that the tip radius R remained constant throughout the measurements, since F_{AD} should rapidly increase with R according to eq 3.

3. RESULTS AND DISCUSSION

3.1. Computation of Experimental Errors via the Standard Error.

A full set of experimental raw data (light blue circles) collected continuously for 2−3 h (∼5000 data points) is shown in Figure 2a. The data have been smoothened with the standard rloess (local regression and first degree polynomial model assigning low weight to outliers) function of Matlab33 with a smoothing coefficient of 0.03 (continuous blue line). The mean of the 5000 data points is shown by the dashed black line. At this point, it is worth defining the main concepts employed in this work, namely, accuracy and precision. For this purpose, we define and employ the metrics accuracy ratio (AR) and interval of error (IE); IE refers to a given radius of an interval throughout.
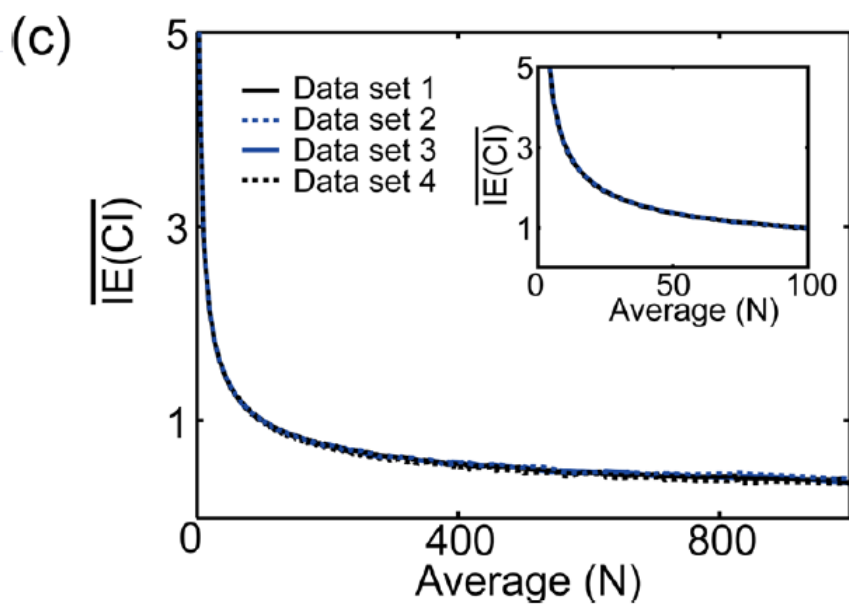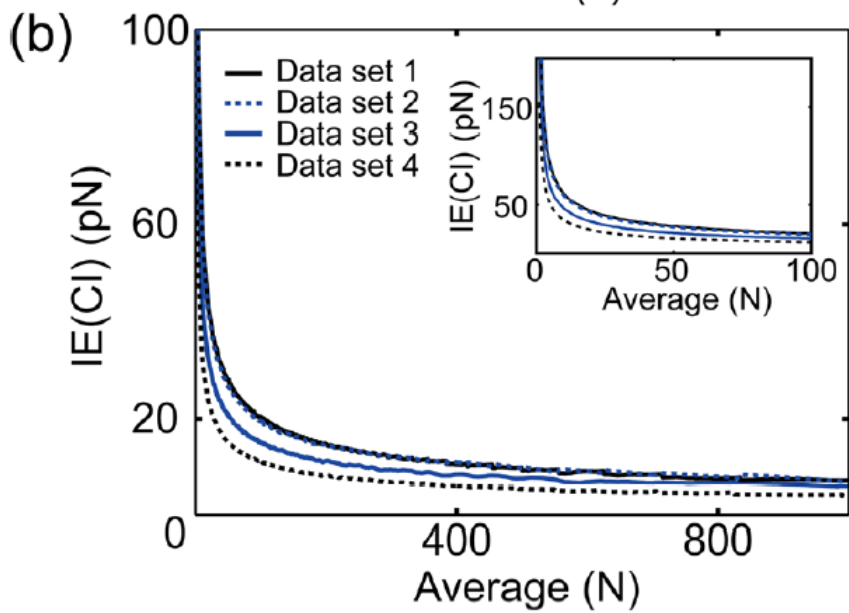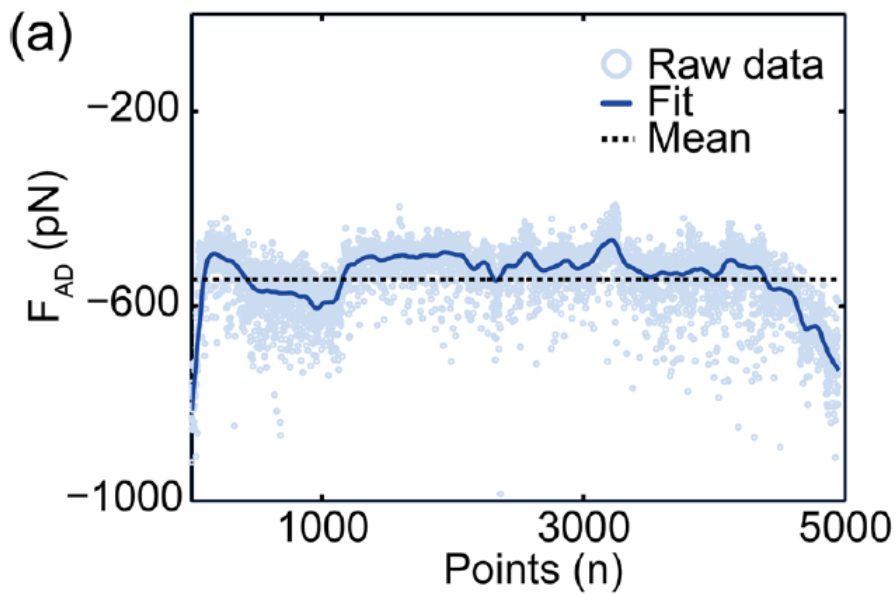
Figure 2. (a) Set of experimental raw data with 5000 data points. The light blue circles are raw data, the black dashed line is the mean of this data set, and the blue line is the smoothened result using Matlab. (b) IE(CI) computed with four different data sets (5000 points each). (c) Normalized IE(CI) of four data sets (at N = 100).

First, accuracy is defined with the concept of AR. AR is employed to compute the confidence that the estimated mean in a measurement, accounting for the error, i.e., an interval of error (IE), as computed with a sample's size N, will include the true mean of the parameter being measured within a given experimental setup. More thoroughly, if we set up an AFM and we do not recalibrate or readjust cantilever-photodiode parameters, i.e., spring constant k, Q factor, etc., during measurements, the constraint AR < 0.05 implies that any IE produced with N data points only will include the true mean of the parameter being measured at least 95% of the times that the measurement is performed. This concept is illustrated with the help of Figure 3 where three intervals IE are shown.
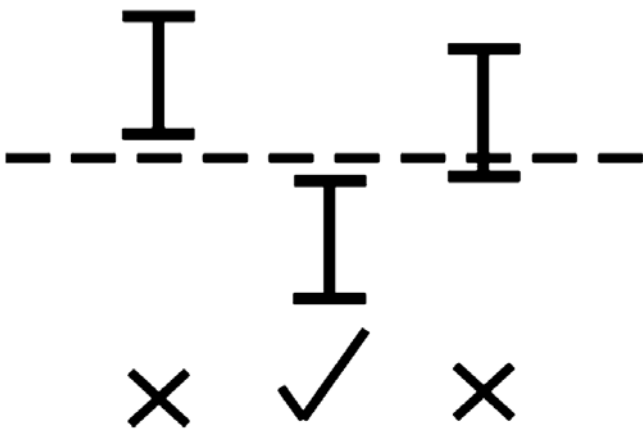

Figure 3. Illustration exemplifying the method to numerically compute the accuracy of the intervals in a measurement.

The first two on the left do not include the true mean (dashed lines), as indicated by crosses, while the other one on the right does (tick). The AR metric in the case of the illustration in Figure 3 produces AR = 2/3 ≈ 0.66, indicating lack of accuracy or a confidence of 33%.

Furthermore, we purposefully neglect the fact that the true mean cannot be measured unless the calibration of the instrument is accurate. This is because, while we acknowledge that systematic errors will be present due to uncertainty in the sensitivity of the laser-photodiode system, k, Q, etc., our focus here is on the detection or identification of the relative contrast necessary to establish heterogeneity in the context of repeatability within an experiment. Thus, when we refer to the true mean, we refer to the mean that would be measured if the measurement was conducted many times, i.e., theoretically infinitely many times, but here we will perform it a minimum of 1000−5000 times (0.5−3 h of continuous measurements), with a given calibration of the system. This greatly simplifies the theory and experimentation while still allowing numerical comparison within experiments, i.e., establishing whether the contrast between two phases on a surface is statistically significant, this being one of the most common ways in which AFM data is employed. In our experiments, we further computed AR in a different way that was at a computationally much higher cost. In particular, we employed combinatorics,[34] with the help of the standard nchoosek function in Matlab, to compute a ratio, i.e., the number of intervals that exclude other intervals divided by the number of possible combinations. The advantage of this method over the method that we employ here is that it does not require assuming a true mean, or computing it. The major disadvantage is computational, but we confirmed (data not shown) that the two methods produced similar results. With this understanding, we assume that 1000−5000 data points are enough to conclude that the true mean coincides with the estimation of the mean with precision above that required in the experiments. Thus, the use of the term true mean should not lead to ambiguity. The IE will then provide information about the precision of the measurement; i.e., in the context of this work, this is the maximum resolution with which two surfaces, properties or phases on a surface, materials, etc., can be distinguished in an experiment. That is, IE is the margin (or interval) of error.

To a first approximation, let us assume that a Student's tdistribution can be employed to define IE via confidence intervals CI (proportional to the confidence level $1 - \alpha$, and as a function of the samples' size N). We term these intervals IE(CI), and these stand for IE as determined from a confidence interval CI. From this, we can write that F_{AD} is the mean of the N data points $\langle F_{AD} \rangle$ with a given uncertainty computed as

$$F_{AD} = \langle F_{AD} \rangle \pm t_{\alpha/2} \frac{\sigma_F}{\sqrt{N}}$$

(5)

where the term in brackets is the sample's mean $\langle F_{AD} \rangle$, σ is the estimate of the standard deviation, N is the sample's size, and $t_{\alpha/2}$ comes from a Student's t-distribution for a given α (here 0.02). The error or precision is then defined by

$$IE(CI, N) = t_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

(6)

and coincides with the product between the 0.98 quantile of the Student's t-distribution and the standard error (N − 1 degrees of freedom). Data acquired on a sapphire's surface has been employed to compute IE(CI) with the use of eq 6 and for four different data sets (5000 data points each), as shown in Figure 2b. The vertical axis corresponds to the IE(CI) at $\alpha = 0.02$, and the horizontal axis is the sample's size N. Reproducibility is also tested here in a weak form, since a single AFM (Cypher AFM from Asylum Research) has been employed in a single lab (LENS at Masdar Institute, UAE). Otherwise, reproducibility has been tested by acquiring data in two ways. First, by not readjusting the system, i.e., the cantilever-photodiode system was not readjusted. The results are shown with the use of continuous black lines (Figure 2b). The data was obtained with a given cantilever and by user A, continuously and for 2−3 h. Second, the data in dashed blue lines were obtained by stopping the collection of data by user A for several minutes and subsequently acquiring data continuously for another 2−3 h (also by user A). As observed in the figure, differences in the magnitude of IE(CI) with increasing N for these two data sets are minimal, i.e., smaller than 1−5%. Third, a different setup was employed by changing the cantilever and making suitable readjustments to the cantilever-photodiode system as it is typical in AFM experiments. In particular, the data in continuous blue lines were obtained by user B by employing a different cantilever from that employed by user A. The data in dashed black lines were obtained by user B with a third cantilever. A general outcome is that the IE(CI) consistently and monotonically decreases with increasing N independently of the user and cantilever throughout experiments. Practically, Figure 2 and expressions 5 and 6 imply that it is possible to increase precision by increasing the samples' size N, as expected from the standard

theory. Furthermore, from these four data sets, it could be deduced that the largest variations in the margin of error, i.e., IE, occurring from experiment to experiment were due to changing cantilevers and users, i.e., with possible variations in tip radius and systematic errors related to the calibration of the spring constant k, Q factor, etc., and readjusting the photodiode. Variations in the standard deviation σ as a function of N however were independent of experiment, setup, user, and cantilever adjustment, as shown in Figure 2c by normalizing IE(CI) at N = 100 as

$$\overline{IE(CI)} = \frac{t_{\alpha/2} \dfrac{\sigma(N)}{\sqrt{N}}}{t_{\alpha/2} \dfrac{\sigma(100)}{\sqrt{100}}}$$

(7)

The results shown in Figure 2c strongly support the assumption that 5000 data points suffice to reach high precision in terms of the population parameters, i.e., mean, standard deviation, etc., and that these values are representative of the system under study and underlying population.

3.2. Testing the Validity of the Standard Error as a Parameter for Comparing Data Sets.

In order to establish the validity of the above standard method to compare between materials, the IE(CI) metric, as computed via eq 6, has been put to test by employing a sapphire's surface (and four data sets of 5000 data points each) as a model sample which is here assumed to be homogeneous in terms of F_{AD}. For this purpose, the data from Figure 2 (user B) has been grouped into sets of N = 30 data points (Figure 4a) and means (black dots) and IE(CI)'s (continuous blue lines) have been computed via eqs 5 and 6.
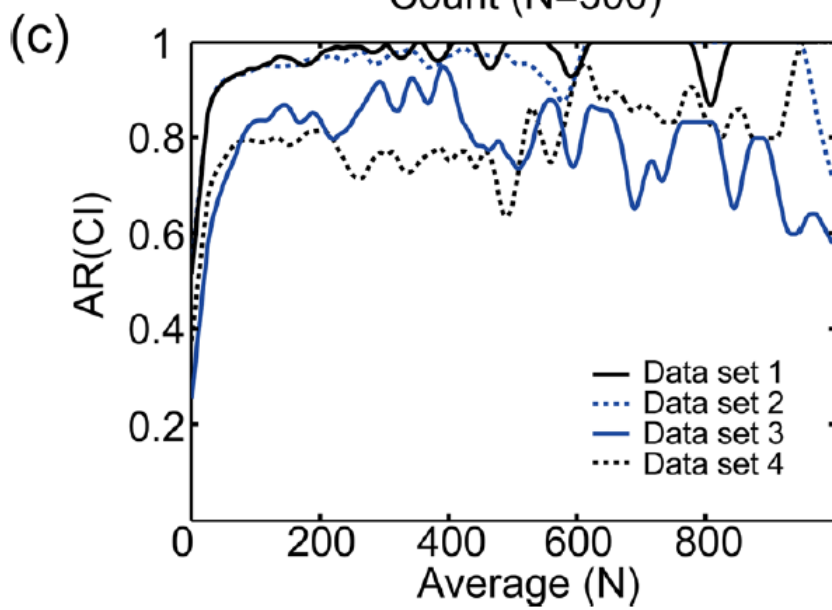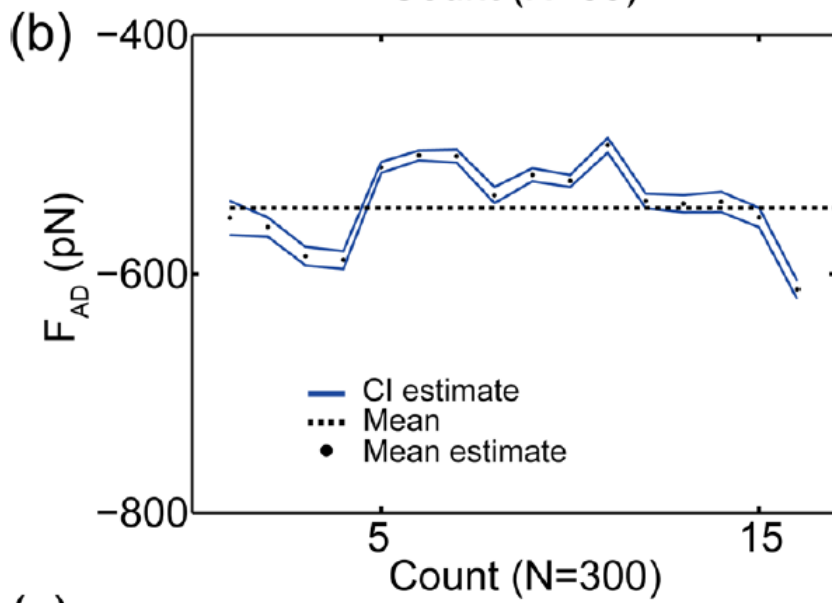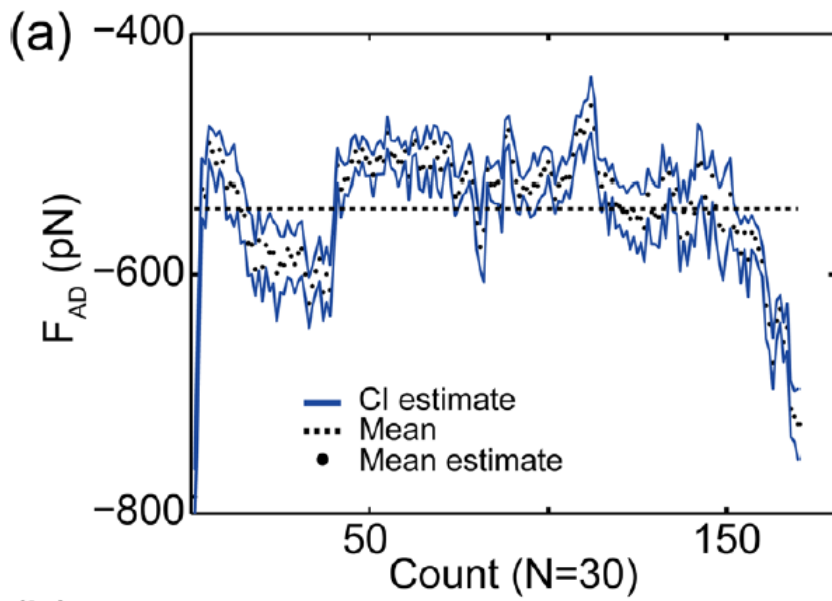
(a)

(b)

(c)

Figure 4. (a) IE(CI)'s calculated from user B's data with N = 30. The black dots are means for N = 30, the black dashed line is the mean for N = 5000, and the blue lines are IE(IC)'s. (b) IE(CI)'s calculated from user B's data with N = 300. The black dots are means for N = 300, the black dashed line is the mean for N = 5000, and the blue lines are IE(IC)'s. (c) The behavior of AR(CI)'s of four data sets with respect to sample size.

The mean of the 5000 data points is shown with the help of black dashed lines, as previously done in Figure 2a. It can be readily concluded by inspection that most of the IE(CI)'s do not include the mean calculated with the 5000 data points. This situation does not improve by increasing the number of data points to N = 300 (Figure 4b). The immediate practical implication deduced from Figure 4 is that, by recording sets of N = 30, or even N = 300 data points, and by computing intervals of error IE(CI) employing eq 6, it would be erroneously concluded that the sapphire surface presents nanoscale heterogeneity in terms of F_{AD}. This conclusion is erroneous, since the sapphire sample is to be employed here as a model system to establish nanoscale homogeneity. More fundamentally, since the data for each data set in Figure 2 was collated on a given physical location of the surface only, the implication is that the adhesion, i.e., mean and error in adhesion, is time dependent. A direct consequence is that the force of adhesion also depends on the number of points that the user takes (since this is time-dependent also). Furthermore, a comparison between two data sets on a given location on a given sample, i.e., here a point on a sapphire's surface, might lead to establishing heterogeneity on that point. It is an objective of this work however to establish consistency between measurements on, at least, a given location of a sample, so comparisons between different locations of a sample, or between different samples, can then be carried out. Thus, this inconsistency needs to be addressed and a different method employed. At this point, it is noted that the normality of the data sets, when averaging over N = 200−300 points and by taking at least M = 10−30 data points from these groups, was established with the use of both the standard jbtest and lillietest normality tests from Matlab. Practically, this implies that means obtained from N = 200−300 data points would have to be employed to construct a normal distribution from which inferences similar to those in eqs 5 and 6 could be made.[20] Nevertheless, the data sets did not pass the test when averaging

over smaller values of N. Thus, in order to get a data set normally distributed with 10−30 data points, approximately 2000−9000 data points would have to be acquired in order to accept or reject the hypothesis of heterogeneity. This condition imposes a great limitation timewise, since it takes at least fractions of a second to acquire each data point, here 2 s per data point implying 1800 data points per hour. More importantly, this could have affected the inconsistent results reached via eq 5 and exemplified with the use of Figure 4. On the other hand, we confirmed that, even when the data sets passed the jbtest and lillietest normality tests, this inconsistency prevailed, i.e., it followed from employing eqs 5 and 6 that the sapphire's surface was different from itself depending on the sample's set in terms of F_{AD} (data not shown). At this point, we did not seek or attempt to draw conclusions about causal connections that could have led to this outcome, and while we understand that this issue might require a more sophisticated analysis, we focused the study in the direction of establishing general criteria to ensure that reproducibility and repeatability, in terms of accuracy and precision, would hold without the need of restricting assumptions.

3.3. Inconsistency of Errors in a Given Data Set and the Accuracy Ratio (AR) Metric Test.

We now define the concept of AR, or inclusion interval, as qualitatively illustrated in Figure 3, and quantify it via a metric AR as follows. From the IE(CI) metric computed from eqs 5 and 6, we define an accuracy ratio AR(CI, N) as

$$AR(CI, N) = \frac{Excluded(IE(CI))}{Total}$$

(8)

where N is the number of points (or sample's size) employed to compute IE(CI), excluded (IE(CI)) makes reference to the number of IE(CI)'s in the experiment that exclude the 5000 data points mean or true mean, and total makes reference to the total number of CIs (here 5000/N). The AR(CI, N) metric has been employed to address how well IE(CI)'s do in including the mean of 5000 data points as a function of N (Figure 4c). The legends and data are the same as those in Figure 2b. The IE(CI)'s do best when including less than 30 data

points, i.e., N < 30. In particular, for the data in Figure 4a, AR(CI, 30) = 0.70 (70% of the intervals exclude the mean), and in Figure 4b, AR(CI, 300) = 0.75 (75% of the intervals exclude the mean). In summary, it can be stated that, in general, the AR(CI, N) metric is too large throughout when employing intervals of error computed with the use of eqs 5 and 6 and therefore not acceptable. Now, we seek to find a method with a metric that provides (1) an error in the form of a radius or interval that becomes smaller, i.e., higher precision, as the sample's size N increases while (2) also providing accuracy, i.e., the intervals should include the true mean. The method (standard error) described in the previous section fails in the second requirement, as exemplified via Figure 4. Thus, we seek a methodology that measures uncertainty more robustly than eqs 3 and 4 and further seek to reduce assumptions regarding underlying distributions and other assumptions regarding independency between the sample's populations. For this purpose, next, we turn to the mean and standard deviation estimates from the sample's populations and construct a theory related to Chebyshev's inequality.[35]
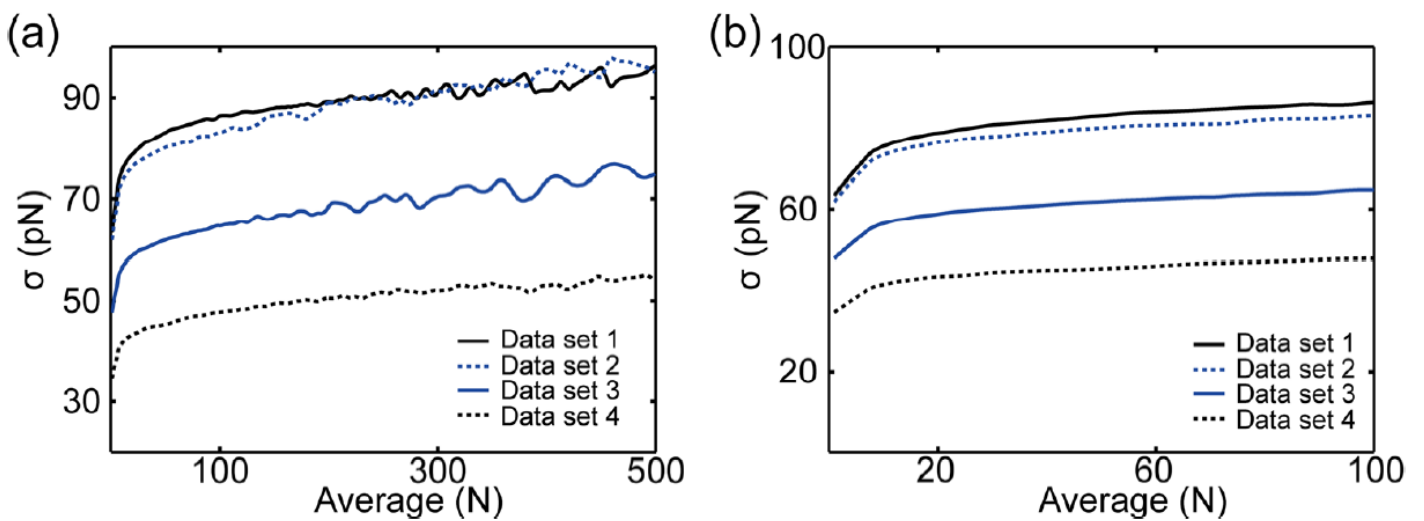

Figure 5. (a) σ calculated from four data sets. (b) σ of four data sets increases slightly as N increases.
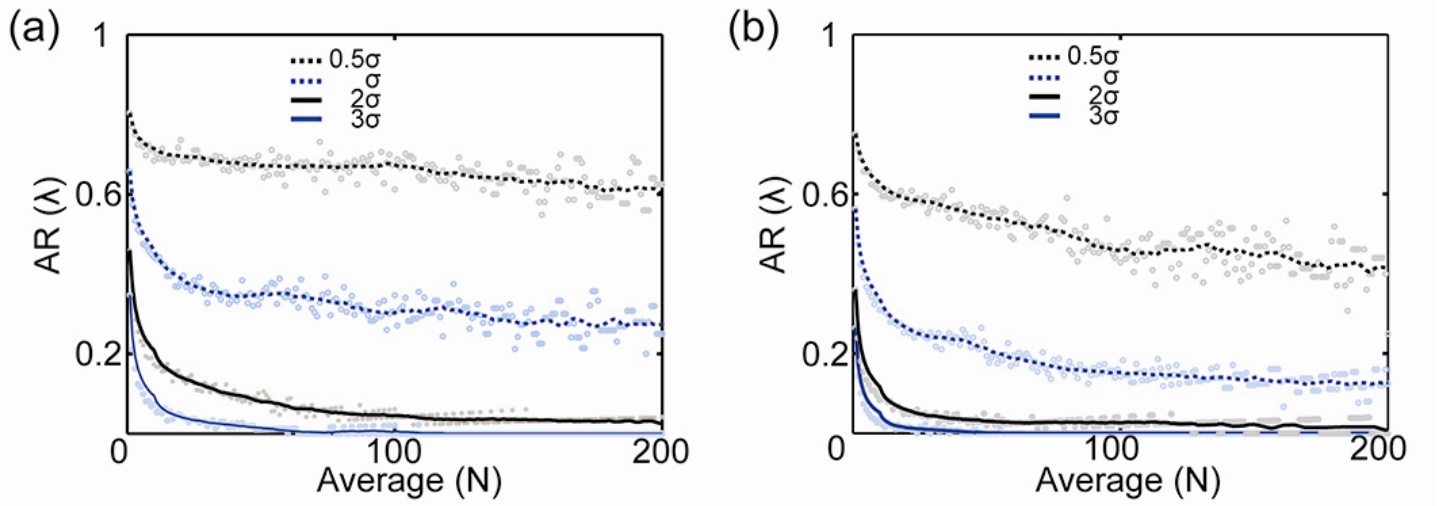
Figure 6. (a) AR(λ)'s behavior with respect to N calculated from user A's data. (b) AR(λ)'s behavior with respect to N calculated from user B's data.

3.4. Definition of Errors Consistent with the Accuracy Ratio (AR) Metric Test.

The same data from Figure 4c has been employed to plot the behavior of the estimates of the standard deviation σ (N − 1 degrees of freedom) as a function of N (Figure 5a). σ increases by ~20% from N = 2 to N ≈ 10, ~10% from N ≈ 10 to N ≈ 100 (Figure 5b), and similarly from N ≈ 100 to N ≈ 500. In general, the magnitude of σ does not even double from N = 2 to N = 500; 2σ(N = 2) < σ(N = 500). Taking the behavior of σ as a function of N into account (Figure 5), we proceed to define values of F_{AD} with errors directly defined from the estimate of σ as

$$F_{AD} = \langle F_{AD} \rangle \pm \lambda \sigma \tag{9}$$

where λ ($\lambda > 0$) is a factor that can be employed to select a given width for the error and can be related to the parameter k in Chebyshev's inequality. Thus, provided N < 500 and considering the results from Figure 5c, halving λ necessarily leads to higher precision. The precision of the measurement is then given in terms of IE as parametrized by λ as

$$IE(\lambda) = \lambda\sigma \qquad (10)$$

The practical use of eqs 9 and 10 now reduces to deducing an appropriate value of λ that is consistent with the measurements. For this purpose, we define an accuracy ratio $AR(\lambda)$ or inclusion interval in terms of the estimate σ directly from the intervals given by $IE(\lambda)$ as

$$AR(\lambda) = \frac{Excluded(IE(\lambda))}{Total} \qquad (11)$$

The above expression quantifies the accuracy of the measurement and can be employed to test the validity of eq 9; i.e., now $IE(\lambda)$ is a function of N in the sense that for a given λ there is a minimum N that will comply with the AR criteria in eq 11. In particular, we would like $AR(\lambda) < 0.05$ with a meaning similar to that of a confidence interval of 95%. That is, in practice, $AR(\lambda) < 0.05$ implies that 95% of the intervals defined by prescribed values of the parameters λ and N via eq 10 will include the true mean. Or in simpler terms, 95% of the times an experimentalist gives an error parametrized by an interval computed via eq 9, i.e., N data points, a prescribed λ and σ as estimated from the N sample's data points, the interval will include the true mean as illustrated by Figure 3. The dependence of $AR(\lambda)$ on N and σ is shown in Figure 6. The vertical axis is $AR(\lambda)$, and the horizontal axis stands for N. The values of λ are 0.5 (dashed black lines), 1 (dashed blue lines), 2 (continuous black lines), and 3 (continuous blue lines). $AR(\lambda)$ monotonically decreases with increasing N independently of λ (not all data shown but we observed this pattern up to $N = 1000$). Nevertheless, it is readily observed that larger values of λ are required for the condition $AR(\lambda) < 0.05$ to apply with a reasonably small number of data points, i.e., $N \sim 10-100$. This implies that higher precision with an accuracy of $AR(\lambda) < 0.05$ requires more data points. These are the properties that we sought out in our methodology. The reproducibility of this behavior is exemplified by plotting the first set of data from user A in Figure 6a and the second set of data from user B in Figure 6b.
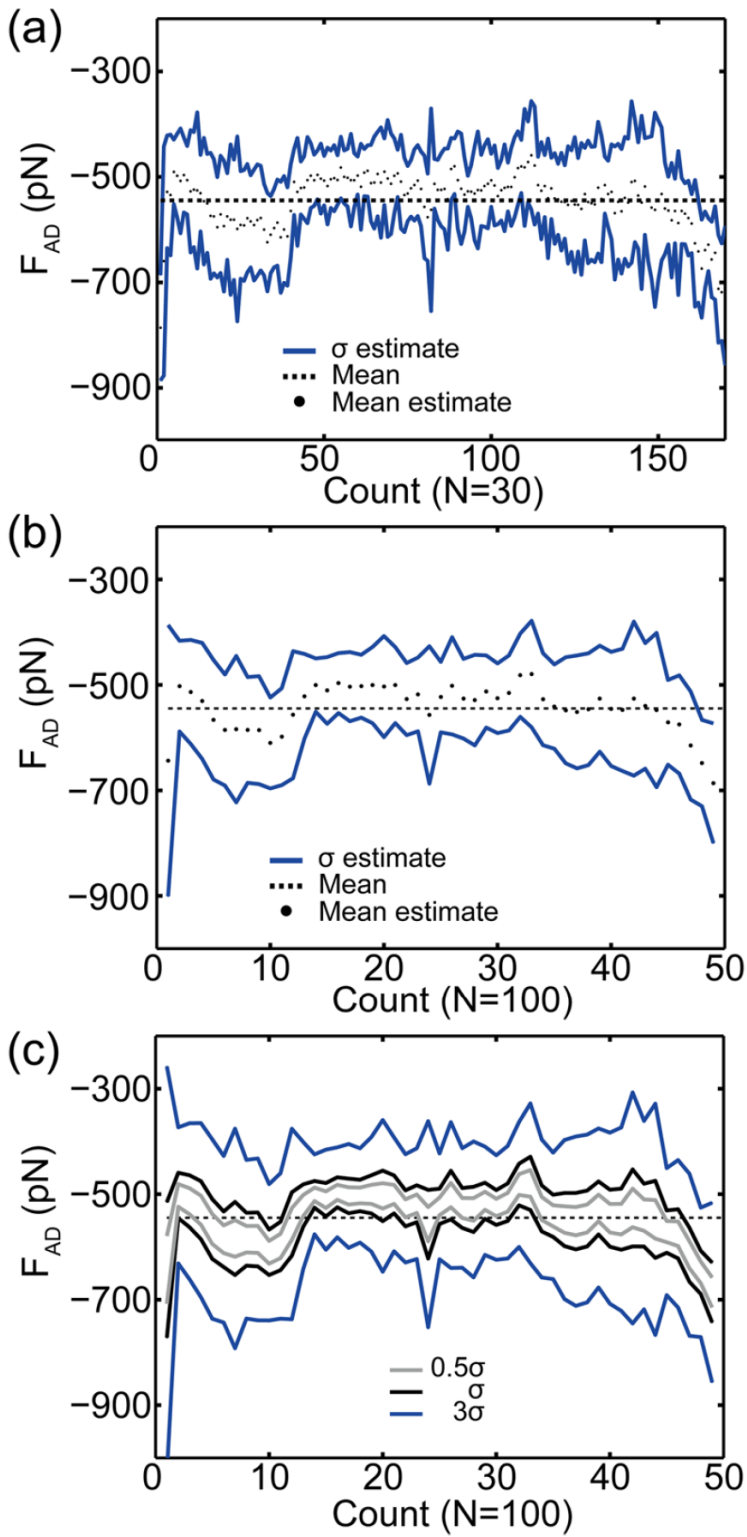
Figure 7. (a) AR(2) with N = 30 calculated from user B's data. (b) AR(2) with N = 300 calculated from user B's data. (c) The behavior of IE(λ)'s calculated with N = 100 under different precision criteria.

An example of how IE(λ) and AR(λ) can provide good estimates for the accuracy and precision of the measurement in AM AFM is given with the help of Figure 7. The data belongs to the second set of 5000 points acquired by user B as in the examples above. We start by establishing accuracy. The actual values of AR in Figure 7a and b are AR(2) = 0.12 (N = 30 in Figure 7a), AR(2) = 0.04 (N = 100 in Figure 7b), and AR(2) = 0.00 (and N = 300 not shown), where λ = 2 throughout and N varies from 30 to 100 and 300, respectively. The accuracy of the measurement is determined by AR(λ) and λ = 2 is already prescribed here, implying that N, i.e., the minimum number of data points required to reach a given accuracy in terms of AR(λ), is the only free parameter. Thus, if an accuracy of 95% or more is required by the experiment, a minimum value of N needs to be found such that AR(2) < 0.05. The pair N = 30 and λ = 2 should then be excluded, since AR(2) > 0.05 (N = 30 in Figure 7a). The pair N = 100 and λ = 2 however is sufficient for this experiment, since AR(2) < 0.05 (N = 100 in Figure 7b). The pair N = 300 and λ = 2 is not appropriate here, since for N = 300 more than 99% of the intervals produced by the experimentalist would lead to including the true mean, and, in this case, only 95% of the intervals are required to include it. That is, N = 300 provides more accuracy than required. Let us now examine the interval of error (precision) in the experiment. Since λ = 2, the estimated error interval is IE(2) = 2σ (N = 100). In particular, in this experiment and for N = 100, the standard deviation is ~48 pN, implying that the total uncertainty in the measurement is ~96 pN. Thus, practically, heterogeneity could be established in this case if the means in F_{AD} of two materials or phases were at least 96 pN apart (employing N = 100 data points). If more precision was required, then a smaller value of λ would be necessary, but this would also necessarily lead to larger sample sizes N in order to maintain a level of accuracy as shown in Figure 6. For the sake of completeness, the behavior of the interval of error IE(λ) or precision with increasing λ is shown in Figure 7c. In this case, N = 100 throughout. The accuracy of the measurement, as quantified by AR(λ), increases with decreasing λ as expected; AR(0.5, 100) = 0.67 (67% of intervals exclude the true mean), AR(1, 100) = 0.31 (31% of intervals exclude the true mean), and AR(3, 100) = 0.00 (0.00% of intervals exclude the true mean), while precision decreases with increasing λ, as can be directly deduced by inspection by noting how the intervals become larger with increasing λ. The main result of the methods presented here can now be written at once: the accuracy of the measurement increases with increasing λ and N, while the precision of the

measurement increases with decreasing λ. The added complication of rejecting or accepting the hypothesis that two materials or phases present nanoscale heterogeneity in a nonarbitrary fashion, for example, by increasing N and decreasing λ arbitrarily, is left as a real experimental example discussed below.

3.5. Comparison between Two Phases: Block Copolymer and Heterogeneity of Phases.

The methodologies derived above are employed next to determine the minimum number of points N to establish heterogeneity with a given maximum interval or margin of error. That is, this exercise exemplifies how heterogeneity can be accepted or rejected in a given scenario with a minimum number of points N. For this purpose, we use a block copolymer polystyrene-b-poly(methyl methacrylate) (PS-b-PMMA) thin film that has been recently shown to exhibit mechanical nanoscale heterogeneity in terms of the Young modulus,[11] i.e., with the use of mechanical contact forces. Amplitude and phase distance curves (APDs) were recorded in the two ordered nanoscale structures (PS and PMMA) in order to reconstruct the force as in the above experiments. The two phases (PS and PMMA) were identified (cross for phase 1 and triangle for phase 2) by inspection of the image scanned in the repulsive regime prior to acquiring the amplitude and phase distance data, since high contrast has been reported in mechanical properties only.[11] The characteristic cylinders of the PMMA can be seen in Figure 8 and have diameters in the order of 10−20 nm as reported elsewhere.[11]
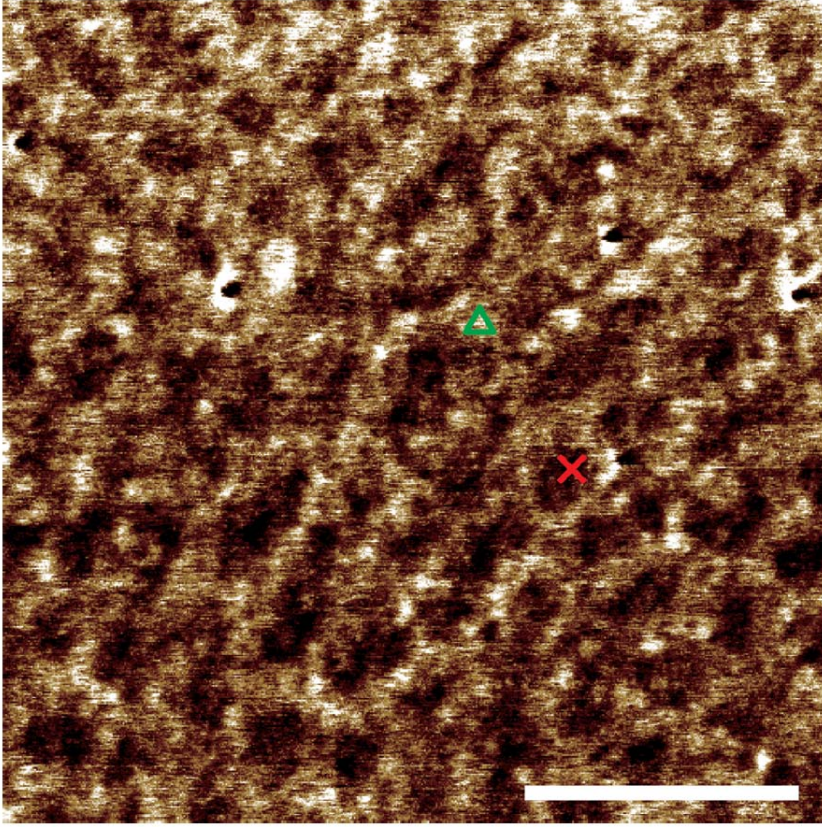
Figure 8. AFM scanning operated in the repulsive regime. The cross indicates phase 1, and the triangle represents phase 2. Scale bar of 100 nm and contrast of 7 degrees in phase from dark brown to white. The red cross indicates the location (darker pixels) from which phase 1 data was acquired, and the green triangle indicates the location (white pixels) from which phase 2 data was acquired.

The aim now is to establish the minimum number of points N necessary to accept the hypothesis that the two phases are different in terms of F_{AD}, i.e., that there is nanoscale heterogeneity in terms of F_{AD}, and provide the margins or intervals of error (radii) with a given confidence level, i.e., 95%. To this end, we define the minimum difference D_m in F_{AD}, accounting for errors, and computed using the metrics eqs 9−11 as follows. Let us first define the difference between estimated means in the two phases as

$$\Delta \mu = \left\| < F_{AD}(1) > - < F_{AD}(2) > \right\|$$

(12)

where (1) and (2) stand for the two phases under comparison and the parameters are identified with the estimates of the means μ1 and μ2, respectively, i.e., $\Delta\mu = \mu2 - \mu1$. If the intervals of error (radii) are computed from eq 10, then the minimum difference between means, accounting for errors, can be defined as D_m and written as

$$D_m = \Delta\mu - 2\lambda\sigma^*$$

(13)

where σ^* is the mean of the two estimates (phase 1 and phase 2 respectively) of the standard deviations. We recall that we know from experimental evidence (Figure 6) that, while σ monotonically increases with N, the rate of change is so small in practice, i.e., $\sigma(N=10) < 2\sigma(N=500)$, that σ can be considered a constant in the context of eqs 10 and 13. A constraint to determine nanoscale heterogeneity based on the computation of D_m can now be written at once as

$$D_m > 0$$

(14)

or

$$\frac{D_m}{\sigma^*} > 0$$

(15)

The advantage of eq 15 is that the minimum difference between means is given in terms of standard deviations. From eqs 12 and 13 the minimum value of λ to establish nanoscale heterogeneity can also be estimated, provided the difference $\Delta\mu$ and the standard deviations are known, by rewriting eq 13 as

$$\frac{\Delta\mu}{2\sigma^*} > \lambda$$

(16)

from which the critical value of λ_c can be written as

$$\lambda_c = \frac{\Delta\mu}{2\sigma^*}$$

(17)

It is worth recalling that λ controls precision (Figure 7c) via eq 10 and this value cannot be selected arbitrarily for a given N, since accuracy (say 95% or AR(λ) < 0.05) also depends on it from eq 11 (Figure 7). In practical terms, the constraint AR(λ)<0.05 will ensure that estimates in F_{AD} for a given material will be consistent in terms of the repeatability (and reproducibility) of experiments while λ will control the minimum difference between means that can be detected, i.e., precision.

minimum N	λ	AR	mean difference	Rep.1	Rep.2	Rep.3	Rep.4	Rep.5
325	0.3	<0.05	D_m [pN]	159	113			
			D_m/σ	2.5	2.5			
250	0.5	<0.05	D_m [pN]	158	113			
			D_m/σ	2.4	1.9			
80	1	<0.05	D_m [pN]	141	117	80	38	85
			D_m/σ	1.6	2.3	1.3	0.7	1.6
12	2	<0.05	D_m [pN]	-17	0	5	170	78
			D_m/σ	-0.5	0	0.1	3.6	1.5
4	3	<0.05	D_m [pN]	-37	-138	-90	-78	-71
			D_m/σ	-1.2	-3.6	-2.1	-1.8	-1.8

Table 1. Numerical Values for the Minimum Number of Data Points N to Establish Nanoscale Heterogeneity from F_{AD} (on PSb- PMMA)[a]

[a]Positive D_m/σ values are required to establish nanoscale heterogeneity, and the minimum number N follows for a maximum value of λ, i.e., N ≈ 80 and λ ≈ 1. Expression 17 predicts a maximum λ_c ≈ 1.5.

The values of D_m (eq 14) and D_m/σ^* (eq 15) obtained for the PS-b-PMMA copolymer (where PS and PMMA are the two phases under comparison) are given in Table 1 (D_m is given in pN) as a function of λ and N while ensuring AR(λ) < 0.05. Approximately 1000 data points were collected for each phase. Note that a minimum of N = 250 data points are required for an accuracy of AR < 0.05 when λ = 0.5. It follows from the table that the minimum number of points to ensure that D_m > 0 is N ≈ 80 (80 points per phase) with λ = 1, since larger values of λ (or intervals of error) already produce negative values for D_m (see Table 1 for λ = 2). Replicates (Rep) are given in Table 1 for λ = 0.3, 0.5, 1, 2, and 3. This value of λ could have been estimated from eq 17 by

calculating the mean from the 1000 data points per phase and the respective standard deviations. In particular, the values were $\Delta\mu = 181$ pN and $\sigma^* = 62$ pN, producing $\lambda_c = 1.5$ (from 1000 data points per phase) which is in agreement with Table 1. These results also imply that, with $N < 100$, differences in the order of $10-10^2$ pN can be detected, i.e., margins of error of $\sim 1-2\sigma$, where $\sigma \approx 30-60$ pN.

For completeness, we next employ the W_{AD} in the same PS−PMMA phases and from the same data. The results are shown in Table 2. The values for D_m become negative for $\lambda > 1$, implying that the minimum number of points N per phase in order to accept that the phases present heterogeneity in terms of W_{AD} is $N \approx 210$ (about 7 min of data acquisition per phase for a scan rate of 0.5 Hz). Then, W_{AD} is a worse parameter to establish nanoscale heterogeneity in terms of the minimum number of points N than F_{AD} for the PS−PMMA pair, since $N \approx 80$ when employing F_{AD} (less than 3 min of data acquisition per phase for a scan rate of 0.5 Hz). The value λ_c could have also been estimated from eq 15; $\Delta\mu = 6.4$ eV and $\sigma^* = 2.5$ eV, producing $\lambda_c = 1.2$ also in agreement with Table 2.

| minimum N	λ	AR	mean difference	Rep.1	Rep.2	Rep.3	Rep.4	Rep.5
210	1	<0.05	D_m [eV]	6	2	2		
			D_m/σ	1.8	0.9	1.6		
145	2	<0.05	D_m [eV]	7	-2	2	-3	
			D_m/σ	3	-0.9	1.3	-1.7	
45	4	<0.05	D_m [eV]	10	-7	0	-2	-7
			D_m/σ	10.5	-2.8	0.4	-1.4	-4.7
26	6	<0.05	D_m [eV]	5	5	-5	-2	-4
			D_m/σ	4.1	4.3	-4.4	-1.8	-3.8

Table 2. Numerical Values for the Minimum Number of Data Points N to Establish Nanoscale Heterogeneity from W_{AD} (on PSb- PMMA)[a]

[a]Positive D_m/σ values are required to establish nanoscale heterogeneity, and the minimum number N follows for a maximum value of λ, i.e., $N \approx 210$ and $\lambda \approx 1$. Expression 17 predicts a maximum $\lambda_c \approx 1.2$.

3.6. Remarks on Establishing Heterogeneity and Interpretation of Metrics.

Here we employ the example in section 3.5 (Figure 8 and Tables 1 and 2) to interpret whether the results from establishing heterogeneity in two locations on a surface, i.e., phases, or two different surfaces, can be interpreted as physically relevant and further provide physical intuition about the metrics. In particular, we focus on the minimum number N to establish heterogeneity and the critical value of lambda λ_c from eq 17. First, the steps to probe heterogeneity involve (1) finding the estimate of the difference in means $\Delta\mu = \mu 2 - \mu 1$ and standard deviations σ^* for a parameter, i.e., adhesion, work of adhesion, etc., for the two different locations on a surface, or two different surfaces, etc. These parameters can be computed from the N experimental points as standard arithmetic means and deviations, respectively. (2) Finding the maximum value of λ necessary to obtain intervals of error computed via eq 10, i.e., $IE(\lambda) = \lambda\sigma$, that satisfy satisfy $D_m > 0$ in eq 13. Provided $D_m > 0$ is satisfied, heterogeneity is established within the given error $IE(\lambda) = \lambda\sigma$. We note that the value λ further needs to satisfy $AR(\lambda) < 0.05$ (0.05 for confidence of 95%) in eq 11.

The criterion $AR(\lambda) < 0.05$ will allow establishing the minimum number of points necessary to employ a given value of λ with which to compute the error $IE(\lambda) = \lambda\sigma$. Typically, the smaller the λ, the more points N are required to satisfy $AR(\lambda) < 0.05$ (see Figure 6). It could be argued that a sufficiently large number of points N can always be employed to satisfy $AR(\lambda) < 0.05$, since AR monotonically decreases for any given λ (see Figure 6). Therefore, heterogeneity can always be established by sufficiently increasing the number of points N in the measurement. Ambiguity in terms of whether the established heterogeneity is physically meaningful however can be removed by considering the maximum value of λ found in step 2. In particular, we showed in the previous section that this maximum value of λ can be approximated via $\lambda_c \approx \Delta\mu/(2\sigma^*)$ (eq 17) where estimates for all parameters are computed from the sampled points. Physically, large values of λ_c, i.e., $\lambda_c \gg 1$, imply a large heterogeneity in the parameter being exploited, i.e., force of adhesion, etc., whereas smaller values of λ_c, i.e., $\lambda_c < 1$, imply a limited relevance of the heterogeneity found in that parameter. An example of a large value of λ_c (large heterogeneity) is illustrated in Figure 9a, and an example of a small value of λ_c (reduced heterogeneity) is found in Figure 9b.
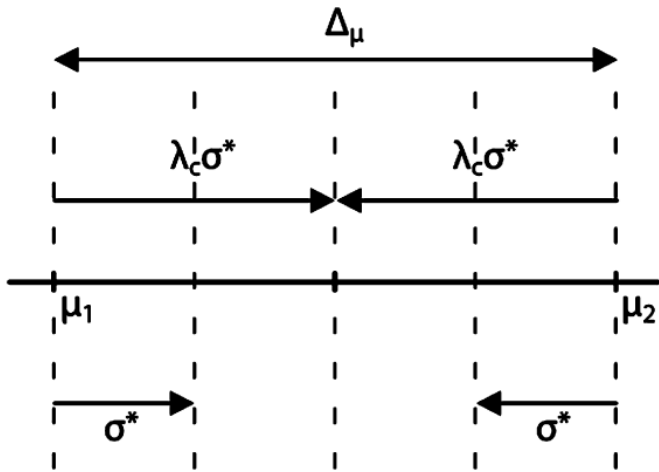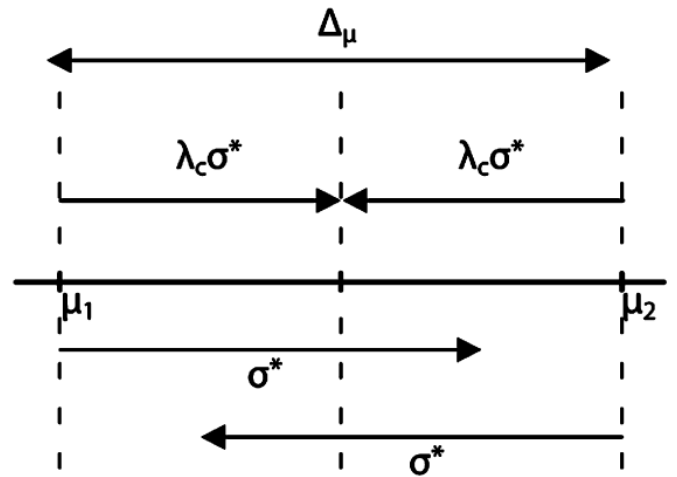
Figure 9. Example of comparisons between two phases for (a) a case where $\lambda_c \gg 1$ and (b) a case where $\lambda_c < 1$.

Then, one could state that one or another parameter, i.e., F_{AD}, W_{AD}, etc., is best by invoking either a minimum number of points N or a larger critical lambda λ_c value required to establish heterogeneity (see Figure 9). In the case presented in section 3.5, i.e., the copolymer, and Figure 8 and Tables 1 and 2, F_{AD} is better than W_{AD} in both counts ($\lambda_c \approx 1.5$ (N = 80) and $\lambda_c \approx 1.2$ (N = 210) for F_{AD} and W_{AD}, respectively). Finally, a loose interpretation might be appropriate here in terms of the minimum number N corresponding to the λ_c parameter. First, it seems plausible that a large number N will be required for a given λ_c whenever there are time-dependent variations in the underlying distributions. This is because otherwise the data points would always lie in the vicinity of a given static mean. In the cases that we have explored, however, this does not occur. See, for example, the seemingly discrete jumps in the value of F_{AD} in Figure 2a. The concept of λ_c should, in any case, and because of the definition of AR in eq 11, apply for any underlying distribution. Second, a small value of λ_c does not necessarily involve a large minimum number of points N or viceversa. That is, from eq 17, λ_c depends on how far apart the means are relative to the standard variations. Then, provided $\Delta\mu \gg \sigma^*$, λ_c will always be large, i.e., $\lambda_c \gg 1$. This is a very desirable result, since if the parameter λ_c is to be employed to determine how different two surfaces, pixels, or phases are, i.e., large heterogeneity, the difference in the estimates of the means $\Delta\mu$ should be very large compared to σ^* and λ_c should increase with this difference as in the illustration in Figure 9.

4. CONCLUSIONS

In summary, we have discussed the conditions for which data extracted from force measurements can be employed to determine whether a surface presents nanoscale compositional heterogeneity. The key parameters are the number of data points and the accuracy and precision that can be achieved by employing N data points only. First, we have shown that reporting the sample's mean, standard errors-computed by assuming normally distributed populations and/or other assumptions such as independent and identically distributed distributions, etc.-and standard deviations only might lead to inconsistent conclusions. Then, a set of metrics have been introduced in terms of accuracy and precision in the measurements that have been shown to ensure reproducibility and repeatability in experiments. These metrics have been employed to conclusively establish the presence or absence of compositional heterogeneity via a given parameter derived from force measurements with a given number of data points N, and with a given margin of error while ensuring that the results are repeatable and reproducible. In practical terms, we have shown how a parameter or material property of choice can be compared from phase to phase, surface to surface, or pixel to pixel while unambiguously and robustly establishing the absence or presence of heterogeneity in terms of that parameter or material property. We have also shown that the exploitation of alternative parameters is interesting from an experimental point of view, since a material might be highly heterogeneous in terms of a given property but homogeneous in terms of another. Furthermore, the proper selection of the parameter might lead to the determination of heterogeneity with fewer data points. These arguments reinforce the relevance of multiparametric models and analysis in the field of nanoscale force spectroscopy. Finally, it could be arguably stated that the main complication in this work related to dealing with the uncertainties and variations within a given experiment, phase of an otherwise homogeneous material. We have not given an interpretation for the nature of such phenomena and have not attempted to discuss causal connections in this work. Instead, we have based the study on forcing a given material or phase to quantitatively behave as homogeneous for the sake of comparison with other materials. We do not ignore, however, that a careful analysis of such phenomena might lead to a deeper understanding of the material's surface and the underlying distributions.[36] Finally, this work should aid to produce robust

comparisons between data sets originating from nanoscale force measurements and will assist to produce repeatable and reproducible outcomes in the field.

ACKNOWLEDGMENTS

REFERENCES

(1) Brydson, R. M.; Hammond, C.; Mowbray, D.; Gibbs, M. R. J.;Todd, I.; Grell, M.; Hamley, I. W.; Geoghegan, M.; Jones, R. A. L.;Leggett, G. J. Nanoscale Science and Technology; Wiley: Chichester, U.K., 2005.

(2) Souier, T.; Santos, S.; Al Ghaferi, A.; Stefancich, M.; Chiesa, M. Enhanced Electrical Properties of Vertically Aligned Carbon Nanotube-Epoxy Nanocomposites with High Packing Density. Nanoscale Res. Lett. 2012, 7, 630.

(3) Masuda, T.; Matsuki, Y.; Shimoda, T. Spectral Parameters and Hamaker Constants of Silicon Hydride Compounds and Organic Solvents. J. Colloid Interface Sci. 2009, 340, 298−305.

(4) Liu, Y. H.; Wang, D.; Nakajima, K.; Zhang, W.; Hirata, A.; Nishi, T.; Inoue, A.; Chen, M. W. Characterization of Nanoscale Mechanical Heterogeneity in a Metallic Glass by Dynamic Force Microscopy. Phys. Rev. Lett. 2011, 106, 125504−125507. The Journal of Physical Chemistry C Article DOI: 10.1021/acs.jpcc.5b04456 J. Phys. Chem. C 2015, 119, 18267−18277 18276

(5) Tai, K.; Dao, M.; Suresh, S.; Palazoglu, A.; Ortiz, C. Nanoscale Heterogeneity Promotes Energy Dissipation in Bone. Nat. Mater. 2007, 6, 454−462.

(6) Dokukin, M. E.; Guz, N. V.; Gaikwad, R. M.; Woodworth, C. D.; Sokolov, I. Cell Surface as a Fractal: Normal and Cancerous Cervical Cells Demonstrate Different Fractal Behavior of Surface Adhesion Maps at the Nanoscale. Phys. Rev. Lett. 2011, 107, 028101−028104.

(7) Eichler, A.; Moser, J.; Chaste, J.; Zdrojek, M.; Wilson-Rae, I.; Bachtold, A. Nonlinear Damping in Mechanical Resonators Made from Carbon Nanotubes and Graphene. Nat. Nanotechnol. 2011, 6, 339−342.

(8) Tian, B.; Al-Jamal, W. T.; Bossche, J.; Kostarelos, K. Design and Engineering of Multifunctional Quantum Dot-Based Nanoparticles for Simultaneous Therapeutic-Diagnostic Applications. In Multifunctional Nanoparticles for Drug Delivery Applications; Svenson, S., Prud'homme, R. K., Eds.; Springer: New York, 2012; pp 345−365.

(9) Baxamusa, S. H.; Gleason, K. K. Random Copolymer Films with Molecular-Scale Compositional Heterogeneities that Interfere with Protein Adsorption. Adv. Funct. Mater. 2009, 19, 3489−3496.

(10) Rafiee, J.; Mi, X.; Gullapalli, H.; Thomas, A. V.; Yavari, F.; Shi, Y.; Ajayan, P. M.; Koratkar, N. A. Wetting Transparency of Graphene. Nat. Mater. 2012, 11, 217−222.

(11) Herruzo, E. T.; Perrino, A. P.; Garcia, R. Fast Nanomechanical Spectroscopy of Soft Matter. Nat. Commun. 2014, 5, 3126.

(12) Costa, L.; Rodrigues, M. S.; Newman, E.; Zubieta, C.; Chevrier, J.; Comin, F. Imaging Material Properties of Biological Samples with a Force Feedback Microscope. J. Mol. Recognit. 2013, 26, 689−693.

(13) Cartagena, A.; Raman, A. Local Viscoelastic Properties of Live Cells Investigated Using Dynamic and Quasi-Static Atomic Force Microscopy Methods. Biophys. J. 2014, 106, 1033−1043.

(14) Voss, A.; Dietz, C.; Stocker, A.; Stark, R. W. Quantitative Measurement of the Mechanical Properties of Human Antibodies with Sub-10-Nm Resolution in a Liquid Environment. Nano Res. 2015, 8, 1987.

(15) Eaton, P. J.; Graham, P.; Smith, J. R.; Smart, J. D.; Nevell, T. G.; Tsibouklis, J. Mapping the Surface Heterogeneity of a Polymer Blend: An Adhesion-Force-Distribution Study Using the Atomic Force Microscope. Langmuir 2000, 16, 7887−7890.

(16) Garcia, R.; Gómez, C. J.; Martinez, N. F.; Patil, S.; Dietz, C.; Magerle, R. Identification of Nanoscale Dissipation Processes by Dynamic Atomic Force Microscopy. Phys. Rev. Lett. 2006, 97, 016103−016104.

(17) Gadelrab, K. R.; Santos, S.; Souier, T.; Chiesa, M. Disentangling Viscous and Hysteretic Components in Dynamic Nanoscale Interactions. J. Phys. D: Appl. Phys. 2012, 45, 012002−012006.

(18) Raman, A.; Trigueros, S.; Cartagena, A.; Stevenson, A. P. Z.; Susilo, M.; Nauman, E.; Contera, S. A. Mapping Nanomechanical Properties of Live Cells Using Multi-Harmonic Atomic Force Microscopy. Nat. Nanotechnol. 2011, 6, 809−814.

(19) Eastman, T.; Zhu, D.-M. Adhesion Forces between Surface-Modified Afm Tips and a Mica Surface. Langmuir 1996, 12, 2859−2862.

(20) Walpole, R. E.; Myers, R. H.; Myers, S. L.; Ye, K. E. Probability and Statistics for Engineers and Scientists; Pearson: Boston, MA, 2011.

(21) Forchheimer, D.; Forchheimer, R.; Haviland, D. B. Improving Image Contrast and Material Discrimination with Nonlinear Response in Bimodal Atomic Force Microscopy. Nat. Commun. 2015, 6, 6270.

(22) Jesse, S.; Vasudevan, R. K.; Collins, L.; Strelcov, E.; Okatan, M. B.; Belianinov, A.; Baddorf, A. P.; Proksch, R.; Kalinin, S. V. Band Excitation in Scanning Probe Microscopy: Recognition and Functional Imaging. Annu. Rev. Phys. Chem. 2014, 65, 519−536.

(23) Sader, J. E.; Uchihashi, T.; Higgins, M. J.; Farrell, A.; Nakayama, Y.; Jarvis, S. P. Quantitative Force Measurements Using Frequency Modulation Atomic Force Microscopy-Theoretical Foundations. Nanotechnology 2005, 16, S94−S101.

(24) Katan, A. J.; van Es, M. H.; Oosterkamp, T. H. Quantitative Force Versus Distance Measurements in Amplitude Modulation Afm: A Novel Force Inversion Technique. Nanotechnology 2009, 20, 165703−165711.

(25) Ramos, J. Tip Radius Preservation for High Resolution Imaging in Amplitude Modulation Atomic Force Microscopy. Appl. Phys. Lett. 2014, 105, 043111−043114.

(26) Vahdat, V.; Grierson, D. S.; Turner, K. T.; Carpick, R. W. Mechanics of Interaction and Atomic-Scale Wear of Amplitude Modulation Atomic Force Microscopy Probes. ACS Nano 2013, 7, 3221−3235.

(27) Israelachvili, J. Intermolecular & Surface Forces, 2nd ed.; Academic Press: San Diego, CA, 1992.

(28) Santos, S.; Guang, L.; Souier, T.; Gadelrab, K. R.; Chiesa, M.; Thomson, N. H. A Method to Provide Rapid in Situ Determination of Tip Radius in Dynamic Atomic Force Microscopy. Rev. Sci. Instrum. 2012, 83, 043707−043717.

(29) Maragliano, C.; Glia, A.; Stefancich, M.; Chiesa, M. Effective AFM Cantilever Tip Size: Methods for in-situ Determination. Meas. Sci. Technol. 2015, 26, 015002.

(30) Amadei, C. A.; Tang, T. C.; Chiesa, M.; Santos, S. The Aging of a Surface and the Evolution of Conservative and Dissipative Nanoscale Interactions. J. Chem. Phys. 2013, 139, 084708.

(31) Yaminsky, V. V. The Hydrophobic Force: The Constant Volume Capillary Approximation. Colloids Surf., A 1999, 159, 181−195.

(32) Santos, S.; Amadei, C. A.; Verdaguer, A.; Chiesa, M. Size Dependent Transitions in Nanoscale Dissipation. J. Phys. Chem. C 2013, 117, 10615−10622.

(33) Matlab and Simulink, R. a. The MathWorks, Inc. Natick Matlab and Simulink, release 2010a; The MathWorks, Inc.: Natick, MA, 2010.

(34) Koh, K. M.; Chen, C. C. Principles and Techniques in Combinatorics; World Scientific Publishing Company: Singapore, 1992.

(35) Papoulis, A. Probability, Random Variables and Stochastic Processes; McGraw-Hill: New York, 1984.

(36) Nguyen, H. K.; Wang, D.; Russell, T. P.; Nakajima, K. Observation of Dynamical Heterogeneities and Their Time Evolution on the Surface of an Amorphous Polymer. Soft Matter 2015, 11, 1425−1433.