

AFM z-piezo calibration method for height measurements in the sub-nanometer to nanometer range using stepped surfaces

Gheorghe Stan

Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg,
MD 20899, USA

Abstract

In this work, we present a method for calibrating height measurements from the z-piezo of an Atomic Force Microscope (AFM) using stepped surfaces with subnanometer staircases. The method analyzes common trace and retrace images obtained from AFM topography scans across straight terraces of these stepped surfaces. By combining these data from a single scan, we can determine both the quantitative height calibration and the linearity of the AFM z-piezo across a full range of heights. This eliminates the need to separately image artifacts of different heights. This approach is significant because it incorporates all measured step heights from a single scan, allowing calibration for any height difference between them. As a result, it serves as a multipoint calibration method that spans a height range from subnanometer levels to tens of nanometers. This contrasts with single-point methods, which typically calibrate against a single height derived from measurements of a single artifact. Our calibration method relies on straightforward metrics, including average measured step heights, standard errors from the mean, relative accuracy, and percentage uncertainty. The most sensitive metrics for detecting potential miscalibration and nonlinearity in the AFM z-piezo are the average residuals and relative accuracy. In contrast, the standard errors and percentage uncertainty primarily relate to the roughness of the stepped surface. We demonstrate the application of this calibration method on scans over a 6H-SiC(0001) surface, with calibrated heights ranging from a minimum step height of 0.756 nm to heights approaching 50 nanometers.

Keywords: step height calibration, subnanometer, nanometer, stepped surfaces, atomic force microscopy

1. Introduction

As various fields such as electronics, biomedicine, and energy conversion continue to scale down from the micrometer to the nanometer range, reliable and straightforward dimensional metrology is essential for the development and improvement of devices and applications. One instrument that provides direct height measurements in the nanometer range is the Atomic Force Microscope (AFM), which uses incremental piezoelectric displacements to track topographical variations at the nanoscale. However, piezo transducers can exhibit nonlinearity and hysteresis responses, as well as creep and aging effects over time. Consequently, like any other metrological instrument, the calibration of an AFM must be periodically verified to ensure accurate and precise height measurements. This calibration is carried out using artifacts whose dimensions are traceable to or referenced from the SI definition of the meter.

Common artifacts used for AFM height calibration include step heights or grooves made from silicon dioxide/silicon gratings through lithographic processes. These calibration gratings are commercially available and feature specific heights ranging from 10 nm to 500 nm. By using these reference heights, the sensitivity of the AFM's z-piezoelectric transducer, measured in nm/V, is determined. Significant efforts have been devoted to developing AFM calibration methodologies based on step heights [1, 2, 3], including a standard procedure detailed in ISO 11952:2019 [4]. It is important to note that these single-step height calibration procedures assess the AFM height calibration only at the height of the specific artifact used, resulting in a unique height for each artifact. Therefore, two or three calibration gratings with lines or grooves of varying heights are used to verify calibration across a range of tens to hundreds of nanometers. This calibration is then extrapolated to cover the entire range of the AFM z-piezo, spanning subnanometers to hundreds of nanometers. However, calibration gratings are not commercially available for heights below 5 nanometers [5], where dimensional measurements are becoming increasingly important in emerging fields such as atomic and nanoscale electronics, 2D materials, and nanoparticles. Various groups have proposed the fabrication of SiO_2 and Al_2O_3 gratings with nanometer-scale heights [5, 6, 7], but these have not yet been developed into commercial products. Therefore, it is essential to establish readily available reference heights to ensure accurate z-calibration and linearity in AFM measurements across the nanometer to subnanometer range.

Viable references for the vertical AFM dimensional metrology at the subnanometer scale can be obtained from the step heights between crystallographic planes. These step terraces can be created either through annealing, such as the annealing under ultra-high vacuum of silicon (Si) surfaces [8], or through hydrogen etching of vicinal surfaces, which produce stepped surfaces in materials like silicon carbide (SiC) [8, 10, 11], aluminum oxide (Al_2O_3) [12], and others. In this context, the metrological reference is given by the natural distance between crystallographic planes, which is intrinsically linked to the material's crystallographic structure.

The heights of monoatomic steps on Si crystallographic surfaces provide a direct, traceable route to the lattice constants of silicon. Over the years, significant effort has been devoted to

establishing an AFM calibration procedure based on Si steps [1, 2, 3, 7, 13]. Notably, as part of the project to determine Avogadro’s number and develop a secondary mass standard with silicon spheres, the silicon lattice spacing $d_{222} = 192.015$ pm [1, 14] was measured through X-ray interferometry. This can then be used to infer a height of 313.560 pm [2] for monatomic steps on Si(111) surfaces, which can be produced through thermal annealing under high vacuum. The most common silicon step configurations used for AFM calibration are staircase and amphitheater structures [15]. In both configurations, the terraces are at least a micrometer wide, allowing for the easy extraction of the step height by measuring the height difference between two adjacent, horizontally leveled terraces. This method aligns with the step height calibration guidelines specified in ISO 5436 [16]. On the other hand, the wide terraces of Si introduce some bowing and nonlinearities that complicate the image analysis, particularly for an AFM with a piezo tube scanner. Regardless of the type of scanner used, accurately restoring the 2D image of an amphitheater configuration to its true topography is challenging. Therefore, a one-dimensional analysis is preferred. Moreover, as the scan size increases, more bowing and nonlinearities are introduced, making large-range calibration by adding consecutive step heights impractical.

A different type of stepped surface can be produced through high-temperature hydrogen etching of surfaces that are slightly miscut at an angle relative to a low-index face, known as vicinal surfaces [8, 10]. This method yields uniformly distributed step-terrace arrays, with terrace widths varying according to the miscut direction and angle. Vicinal surfaces are particularly relevant for semiconductor applications that require heteroepitaxial overlayer growth. One extensively studied material in this context is the hydrogen-etched 6H-SiC(0001), which displays uniformly distributed step-terraces on both the silicon-terminated and carbon-terminated surfaces. The silicon-terminated surface has step heights of 1511.51 pm, corresponding to a full unit cell, whereas the carbon-terminated face features step heights of 755.76 pm, equivalent to a half unit cell [17]. In contrast to Si, SiC step-terraces exhibit greater regularity, enabling control of terrace widths over extended lengths. As a result, within just a few micrometer-square areas, numerous steps can be imaged, revealing varying height differences.

In this study, we used the morphology of 6H-SiC(0001) staircase arrays to establish a set of sequential step-height references for AFM z-calibration, ranging from the smallest step (either a half or a full unit cell) to tens of nanometers in height. This was achieved by analyzing all the steps captured during a scan, allowing us to extract not only a single height difference but also all possible height combinations between the terraces observed in that scan. Figure 1 shows the series of n step heights assembled from the scan of a stepped surface: $h_{\min}, 2h_{\min}, \dots, nh_{\min}$. Thus, we not only verify the z-calibration at multiple points within this range, but we also evaluate the linearity of the piezo transducer throughout the entire z-range.

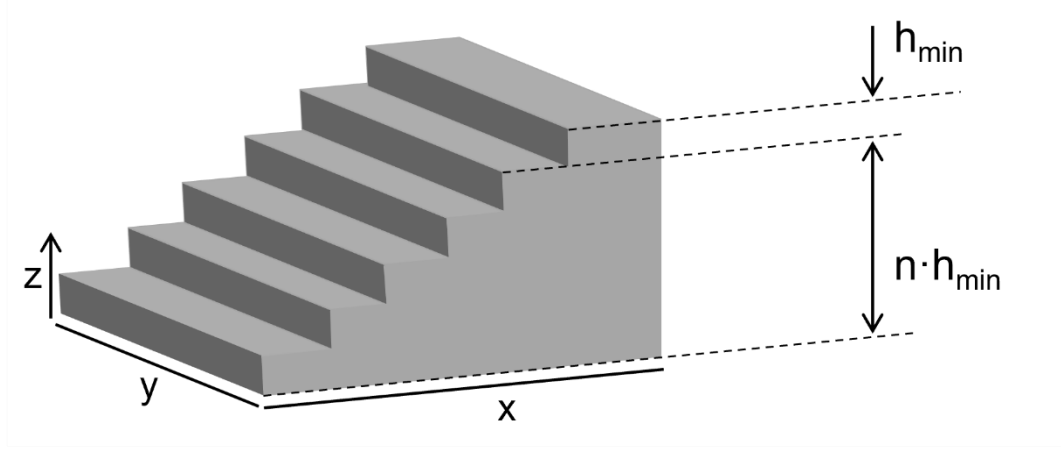


Figure 1. Schematic diagram illustrating the staircase structure of a stepped surface, highlighting the minimum height difference, $h_{\min}$, between consecutive terraces, and the difference in height, $n \cdot h_{\min}$, created by n sequential steps.

A few metrics, including residual differences between theoretical and measured heights, accuracy, and percentage uncertainty, are used to assess the calibration method's performance across all step heights in a single scan. We found that the most sensitive metrics for detecting potential miscalibration and nonlinearity in the AFM z-piezo are the measured average residuals and relative accuracy. In contrast, the standard errors and percentage uncertainty primarily relate to the roughness of the stepped surface.

The calibration method's algorithm is provided in an open-source Python script, which is available in a GitHub repository [18, 19]. The input data consists of 2D arrays of height values sourced directly from the trace and retrace images of a single AFM scan.

2. Method

The input data for the proposed calibration method consists of AFM topography images, as illustrated in Figure 2a. This image represents a $3 \mu\text{m} \times 3 \mu\text{m}$ area of a 6H-SiC(0001) stepped surface, with a nominal step height of 755.76 pm. As is often the case, the AFM image exhibits a noticeable tilt due to the misalignment of the sample surface within the AFM's reference system. The staircase-like arrangement of the height steps remains visible, aligned along the y-axis of the image.

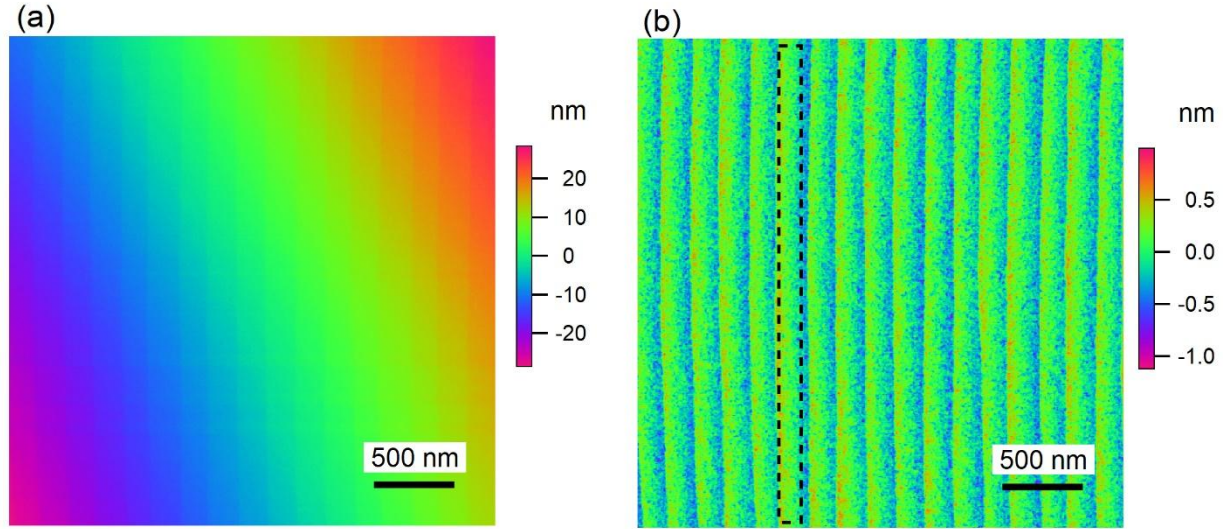


Figure 2. (a) Original AFM topography image (retrace) of a $3\ \mu\text{m} \times 3\ \mu\text{m}$ region of 6H-SiC(0001) stepped surface. (b) Linear plane and line flattened image of the AFM topography shown in (a). The dashed line contour highlights a single terrace of the imaged surface.

The developed algorithm can utilize either the trace image, the retrace image, or both the trace and retrace images from a scan. Each image is typically collected at a resolution of 512 pixels by 512 pixels. Since this calibration is specifically for AFM height measurements, calibrations for the x and y dimensions are not relevant to this discussion and will not be covered.

To achieve a calibration free of artifacts, it is essential to identify a region free of contamination and surface defects. Additionally, it is important to select a region where the surface steps are straight and to align the fast-scan direction perpendicular to these steps. This orientation facilitates a straightforward selection of the flat terraces across all the imaged steps.

The first step in processing the acquired AFM image is to perform a linear plane flattening, followed by a linear line flattening. This helps to eliminate tilt and any potential drift in the slow-scan direction. After these two flattening processes, the resulting AFM image, as shown in Figure 2b, reveals much clearer surface steps for observation. Due to the linear fits along the x-axis, each terrace appears to be locally tilted around the y-axis. To accurately represent the actual topography of the stepped surface, the image needs to be rotated until the terraces are displayed in their proper horizontal orientation. In many commercial AFM analysis software and proposed methods found

in the literature [3], this rotation is performed around a flat region of the surface. Figure 2b highlights one such area located within the flat part of a terrace. While this method is very convenient for image processing, it introduces a residual tilt in the resolved orientation of all the surrounding terraces. This residual tilt becomes increasingly pronounced as the terraces are situated farther from the selected region where the tilt correction is applied. Ultimately, the unresolved tilt of the terraces can negatively impact the accuracy and precision of height and linearity calibrations conducted using the staircase method. In many proposed step height calibration techniques [2, 3], image analysis is often limited to a single step height to mitigate the effects of residual tilt, even when multiple steps of the stepped surfaces are captured in the images.

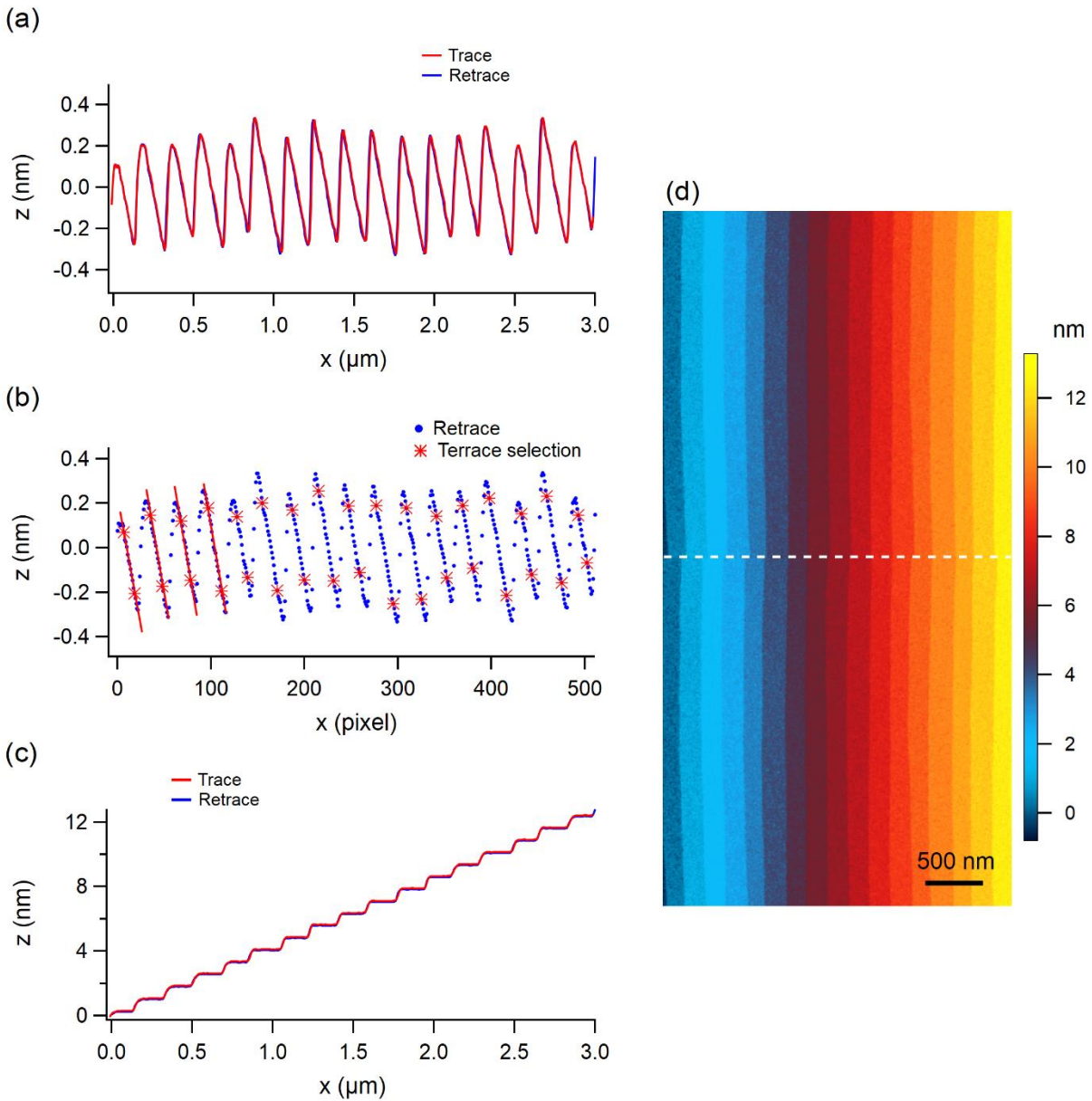


Figure 3. (a) The average profile of the trace and retrace images of the staircase shown in Figure 2b. (b) In the average retrace profile (blue dots), the flat sections of the averaged terraces are marked by red crosses. The red lines connecting these crosses show the linear fits corresponding to the flat part of the terraces. The average slope of these terraces was then used to correct the tilt in both (c) the average trace and retrace profiles and (d) the staircase topography of the imaged 6H-SiC(0001) stepped surface shown in Figure 2. The dashed line in (d) separates the reconstructed trace (bottom) from the retrace (top) images.

To address the issue of residual tilt in our staircase analysis, we propose calculating the average tilt of the scan by incorporating all the imaged steps, on both trace and retrace. First, we obtain the average profiles of the trace and retrace images along the x-axis (fast scanning) by averaging each image in the y-axis (slow scanning) direction. As illustrated in Figure 3a, these average profiles show linear segments across the flat terraces but at a tilted angle, reflecting the post-flattening tilt of the images. Significant deviations from linearity are evident at the ends of the steps, caused by both the edges of the steps and the curvature of the steps along the y-axis of the image. To eliminate these nonlinearities, we manually delineate the extent of the flat terraces on one of the average profiles, either the trace or the retrace profile. In Figure 3b, we demonstrate the selection process for the first four terraces on the left side of the averaged retrace profile. Red crosses have been manually placed at the ends of the flat sections of these terraces. Linear fits are applied between these selected points to determine the slope of each tilted terrace. The average slope of all the terraces in both the trace and retrace tilted profiles is then used to correct for the tilt. This means that the same angle is applied to both the trace and retrace images for tilting correction. The true topography of the stepped surface is illustrated in Figure 3c, which shows the average trace and retrace profiles, and in Figure 3d, which displays the corresponding trace and retrace images. Since both the trace and retrace images will be used in the subsequent analysis, they are combined in Figure 3d, where the white dashed line separates the trace (bottom) from the retrace (top). In both the trace and retrace images, and more clearly in the combined panel shown in Figure 3d, it is evident that the terraces do not have straight, parallel edges in the vertical direction. Consequently, averaging the heights in Figure 3d directly and analyzing them in the resulting histogram could potentially increase the uncertainty in the step height measurements between terraces. To address this issue, we should crop the edge regions and retain only the continuous, flat sections of the terraces in the vertical direction.

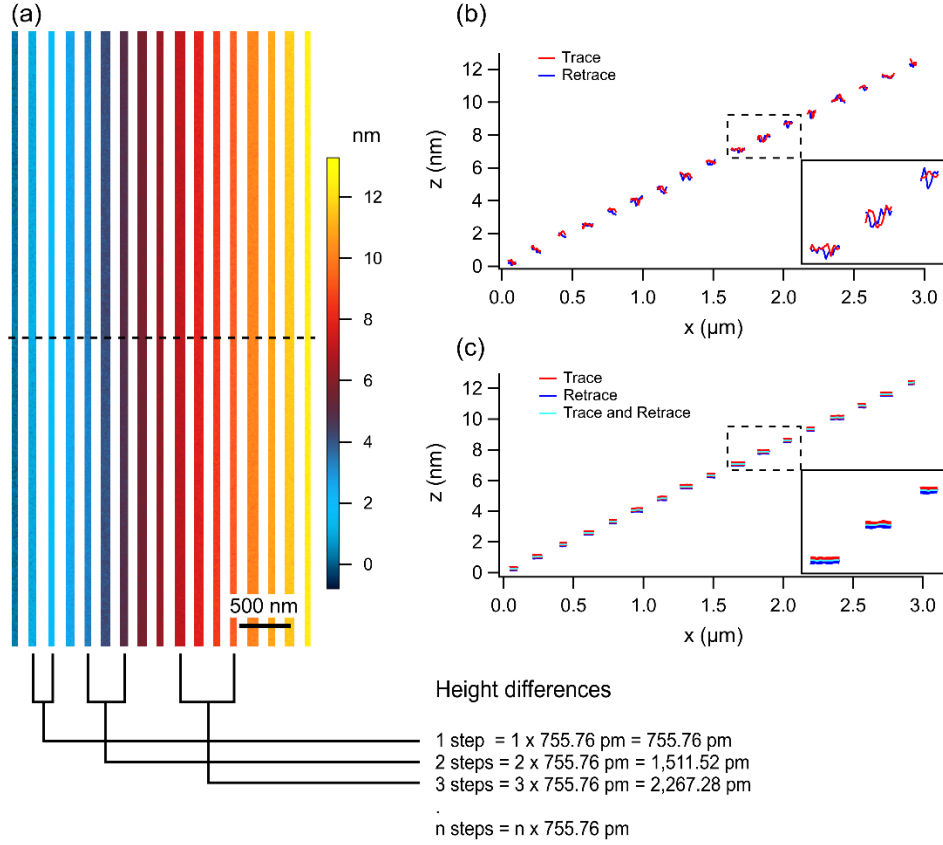


Figure 4. (a) The staircase topography of the 6H-SiC(0001) stepped surface after applying a mask that selects only the flat sections of the terraces along the y-axis. The dashed line separates the trace image (bottom) from the retrace image (top). At the bottom of the panel, combinations of two different terraces illustrate how various height differences are extracted from the data. (b) Individual cross-sections from trace and retrace parts of (a). (c) The average profile of the trace, retrace, and combined trace and retrace data for the flat sections of the terraces shown in (a). The insets in (b) and (c) provide enlarged views of the selected regions from each plot.

The selection of continuously flat regions from terraces in the combined trace and retrace panel in Figure 3d is based on the selection points used for the tilt correction in Figure 3b. Using these points, we define rectangular masks along the y-axis to isolate only the flat regions of terraces. These selected areas are displayed as strips in Figure 4a, with the lowest step height on the left and the highest on the right. By calculating the surface roughness of all the strips in Figure 4a, we found that the average surface roughness is 134 pm for both the trace and retrace images. This surface roughness is highlighted in Figure 4b for individual selected trace and retrace lines from Figure 4a. To measure height differences, we average the profiles along the strips shown in Figure 4a. These average profiles are presented in Figure 4c for trace, retrace, and combined trace and retrace images. The height differences between these average strips will be used for height and linearity calibration across the range from the lowest to the highest

height differences. Including this combined trace/retrace analysis allows us to use both images typically recorded during common AFM scans and to assess the offset between the calibrations based on trace images, retrace images, or a combination of both.

For a given stepped surface with a minimum step height $h_{\min}$, multiples of step heights $s \times h_{\min}$ can be formed by taking successive combinations of the average heights of the considered terraces. Here s counts the possible number of these step heights, ranging from 1 to $N - 1$, where N is the number of the imaged terraces. Given the possible combinations between different terraces, we will have $N - 1$ step heights of $h_{\min}$, $N - 2$ step heights of $2 \times h_{\min}$, $\dots$, $N - s$ step heights of $s \times h_{\min}$, $\dots$, and only one step height of $(N - 1) \times h_{\min}$. A few of these combinations are schematically illustrated at the bottom of Figure 4a. In this instance, there are $N = 17$ terraces within the $3 \mu\text{m} \times 6 \mu\text{m}$ combined trace and retrace image. For two different terraces designated as “High” and “Low”, we calculated their average values, $h_{\text{High}, s}^i$ and $h_{\text{Low}, s}^i$, and their associated standard deviation, $\sigma_{\text{High}, s}^i$ and $\sigma_{\text{Low}, s}^i$, with data from each terrace. This allows us to form a positively defined step height h_s^i

$$h_s^i = h_{\text{High}, s}^i - h_{\text{Low}, s}^i \quad (1)$$

and its associated standard deviation

$$\sigma_s^i = \sqrt{(\sigma_{\text{High}, s}^i)^2 + (\sigma_{\text{Low}, s}^i)^2}. \quad (2)$$

Here the index i goes from 1 to $N - s$ to account for all possible combinations for each step height based on the available terraces.

In Figure 5, we illustrate the algorithm used to combine measurements from the strips shown in Figure 4a. This algorithm calculates the height differences for the smallest height step, which is $h_{\min}$. This is achieved by pairing every two adjacent terraces. Using equations (1) and (2), we calculated the average values and their standard deviations for a height of one $h_{\min}$, ($s = 1$), which are then displayed above the peaks of the histogram in Figure 5. Similarly, the heights from the terraces shown in Figure 4a were used to calculate other step heights. For example, for the step height of $3 \times h_{\min}$ ($s = 3$), pairs of two terraces, with the one in between them skipped, are used. This process continues until the largest step height is determined, which is measured from the lowest terrace (on the left in Figure 4a) to the highest terrace (on the right in Figure 4a).

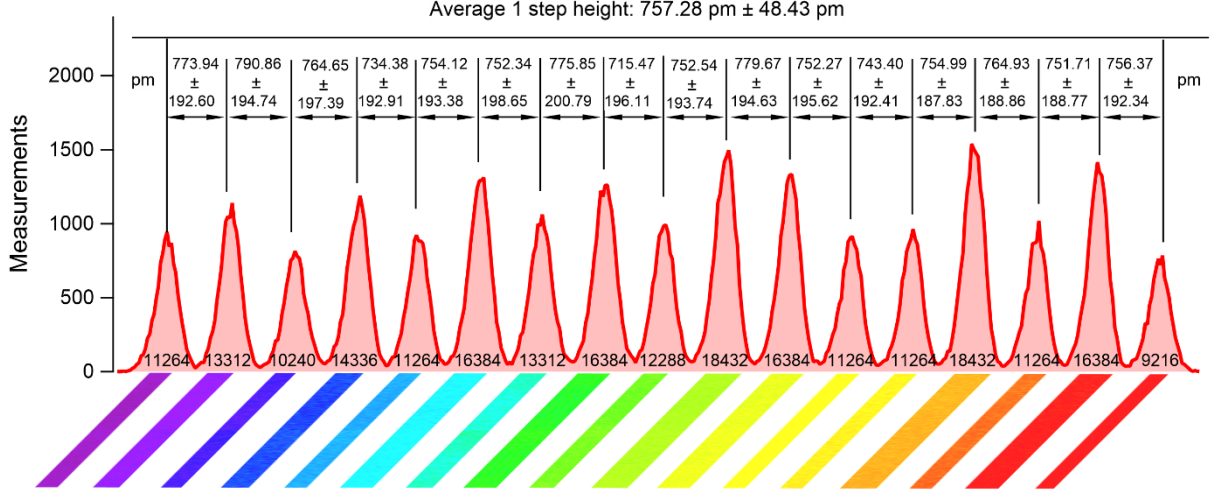


Figure 5. Histogram of the heights of each strip (trace + retrace) from Figure 4a. Each peak in the histogram corresponds to the count of height measurements for each strip. At the bottom, a color gradient visually indicates the variations in height across these terraces. Above the histogram, the differences in height between successive terraces are displayed along with their standard deviations.

Once we have all the step heights calculated for each index s , $s = \overline{1, N-1}$, we use an inverse-variance weighting to calculate the weighted averages

$$h_s = \left(\sum_{i=1}^{N-s} \frac{h_s^i}{(\sigma_s^i)^2} \right) / \left(\sum_{i=1}^{N-s} \frac{1}{(\sigma_s^i)^2} \right) \quad (3)$$

and its associated standard error of the weighted means

$$\sigma_s = \sqrt{1 / \left(\sum_{i=1}^{N-s} \frac{1}{(\sigma_s^i)^2} \right)}, \quad (4)$$

for each of these step heights. As shown at the top of Figure 5, for the step height h_1 , the average and standard error are 757.28 pm and 48.43 pm, respectively. It is also evident that the standard deviation variations across the 16 height differences are minimal. Therefore, the calculated standard error $\sigma_1 = 48.43$ pm given by equation (1) is comparable to the sample standard deviation, which is approximately 190 pm, divided by the square root of the number of samples, 16: $190 \text{ pm} / \sqrt{16} = 48 \text{ pm}$.

The weighted averages and associated standard errors for the step heights across all imaged terraces are presented in Table 1. These values are extracted separately from trace images, retrace images, and their combination. To better illustrate deviations from the actual step height values (column 1 in Table 1), the calculated average step heights are displayed in columns 2, 3, and 4 of Table 1 as residuals, i.e., differences between the measured and theoretical values. This allows us to easily observe the differences between the trace, retrace, and combined trace

and retrace results. For most steps, these differences are within 0.5 pm, and for some, they do not exceed 3.0 pm. With respect to the smallest height difference, $h_{\min} = 755.76$ pm in the first row of Table 1, these variations represent less than 0.4 % of $h_{\min}$. In contrast, a previous step height calibration method [3] for 6H-SiC (0001) revealed an offset as large as 1.5 % for the step height $h_{\min}$ when comparing trace and retrace data. In our measurements, the standard errors for trace and retrace measurements also show small differences from one another (columns 6, 7 of Table 1), all within 1.0 pm. As indicated in Table 1, both the average values and standard errors increase with the step height.

Table 1. Theoretical values, residual (measured average – theoretical) values, and standard errors of the measured step heights of a 6H-SiC(0001) stepped surface, in the range from $h_{\min} = 755.76$ pm to $16 \times h_{\min} = 12,092.20$ pm.

Multiple of the min step, s	Theoretical $s \times h_{\min}$ (pm)	Residual, $h_s - s \times h_{\min}$ (pm), Trace	Residual, $h_s - s \times h_{\min}$ (pm), Retrace	Residual, $h_s - s \times h_{\min}$ (pm), Trace and Retrace	Std error σ_s (pm), Trace	Std error σ_s (pm), Retrace	Std error σ_s (pm), Trace and Retrace
1	755.76	1.32	1.61	1.52	46.95	47.24	48.43
2	1,511.52	2.19	2.02	2.18	48.59	48.87	50.11
3	2,267.28	0.87	0.50	0.75	50.26	50.56	51.85
4	3,023.04	-2.63	-2.48	-2.49	52.22	52.53	53.87
5	3,778.80	-4.91	-4.81	-4.77	54.49	54.78	56.19
6	4,534.56	-5.65	-5.70	-5.63	56.93	57.16	58.66
7	5,290.32	-6.08	-6.50	-6.23	59.55	59.89	61.40
8	6,046.08	-11.87	-11.69	-11.72	62.68	63.05	64.61
9	6,801.84	-13.22	-13.05	-13.10	66.60	66.91	68.58
10	7,557.60	-8.78	-9.30	-9.00	71.04	71.43	73.15
11	8,313.36	-10.31	-10.71	-10.49	76.36	76.98	78.69
12	9,069.12	-10.88	-11.14	-10.97	83.56	84.00	85.92
13	9,824.88	-8.17	-8.43	-8.30	93.90	94.28	96.38
14	10,580.60	4.12	2.23	3.10	108.77	109.21	111.53
15	11,336.40	17.17	14.94	15.99	132.93	133.43	136.09
16	12,092.20	25.48	25.20	25.34	193.40	192.87	196.51

The results shown in Table 1 are illustrated in Figure 6. As in Table 1, we have plotted the residuals rather than the average step heights to enhance visualization.

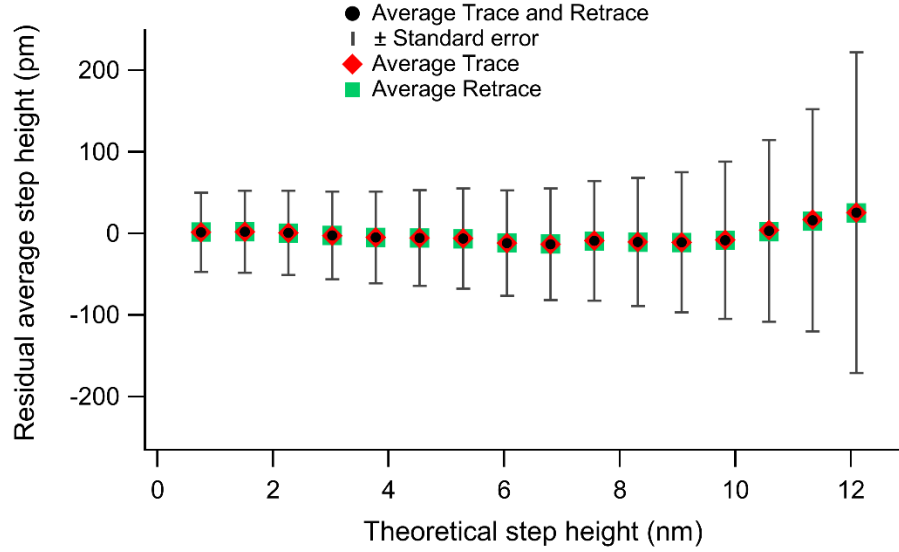


Figure 6. Residual (measured average – theoretical) values (marks) and standard errors (error bars) for step heights between the strips shown in Figure 4a. The average values are shown for each trace, retrace, and combined trace-and-retrace case. The error bars are only shown for the combined trace and retrace, as there are very small variations among these three cases (see columns 6, 7, and 8 in Table 1).

3. Results

We will now use the height measurements (average values and standard errors) to develop a step height calibration procedure. This procedure will provide a quantitative estimate of the difference between the measurements and the actual values, enabling us to apply traceable corrections. To achieve this, we will utilize relative accuracy and percentage uncertainty, defined as

$$\text{relative accuracy} = 100 \% - \frac{|\text{actual value} - \text{measured value}|}{\text{actual value}} \times 100 \% \quad (5)$$

and

$$\text{percentage uncertainty} = \frac{\text{uncertainty}}{\text{measured value}} \times 100 \%. \quad (6)$$

According to the notation in Table 1, for a given multiple s of the minimum step height $h_{\min}$, the actual value is $s \times h_{\min}$, the measured value is h_s , and the uncertainty is the standard error of the mean σ_s . In the following, we will refer to the quantities defined by equations (5) and (6) as the accuracy and uncertainty of the calibration method. We have calculated their values in Table 2 and plotted them in Figure 7, using the data from Table 1.

Table 2. Accuracy (i.e., relative accuracy) and uncertainty (i.e., percentage uncertainty) of the measured step heights from Table 1.

Multiple of the min step, s	Theoretical $s \times h_{\min}$ (pm)	Accuracy (%), Trace	Accuracy (%), Retrace	Accuracy (%), Trace and Retrace	Uncertainty (%), Trace	Uncertainty (%), Retrace	Uncertainty (%), Trace and Retrace
1	755.76	99.82	99.79	99.80	6.20	6.24	6.39
2	1,511.52	99.86	99.87	99.86	3.21	3.23	3.31
3	2,267.28	99.96	99.98	99.97	2.22	2.23	2.29
4	3,023.04	99.91	99.92	99.92	1.73	1.74	1.78
5	3,778.80	99.87	99.87	99.87	1.44	1.45	1.49
6	4,534.56	99.88	99.87	99.88	1.26	1.26	1.30
7	5,290.32	99.89	99.88	99.88	1.13	1.13	1.16
8	6,046.08	99.80	99.81	99.81	1.04	1.045	1.07
9	6,801.84	99.81	99.81	99.81	0.98	0.99	1.01
10	7,557.60	99.88	99.88	99.88	0.94	0.95	0.97
11	8,313.36	99.88	99.87	99.87	0.92	0.93	0.95
12	9,069.12	99.88	99.88	99.88	0.92	0.93	0.95
13	9,824.88	99.92	99.91	99.92	0.96	0.96	0.98
14	10,580.60	99.96	99.98	99.97	1.03	1.03	1.05
15	11,336.40	99.85	99.87	99.86	1.17	1.18	1.20
16	12,092.20	99.79	99.79	99.79	1.60	1.59	1.62

In Table 2, we observe that the accuracy differences between trace data, retrace data, and their combination are all below 0.05 % for all height values. This indicates that the tilt correction applied is optimal for both the trace and retrace data. After this correction, the results remain consistent, whether we use the trace, the retrace, or their combination. Throughout the entire range of calibration heights, the accuracy variations are limited to just a few tenths of a percentage point from 100 %. Furthermore, as shown in Figure 7a, these variations do not exhibit any specific trend related to height. These findings demonstrate excellent linearity for the z-calibration of the AFM within this height range. The percentage uncertainty values displayed in Table 2 (columns 6, 7, and 8) and Figure 7b demonstrate very small variations of only 0.1 % between the trace and retrace data, similar to the accuracy values. Generally, the highest percentage uncertainty is observed for the smallest step height, $h_{\min}$, with a consistent decrease in uncertainty as the step height increases.

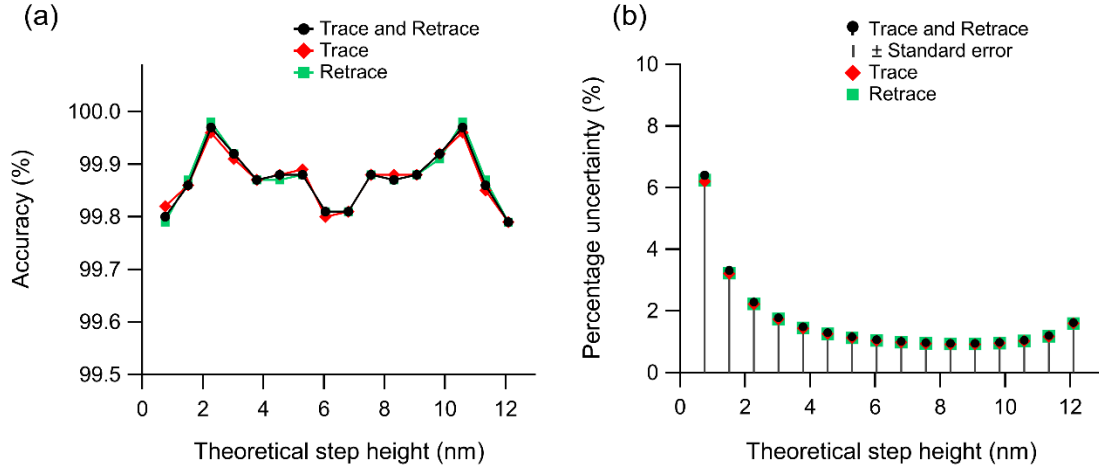


Figure 7. (a) Relative accuracy and (b) percentage uncertainty of the step height calibration applied to the selected strips shown in Figure 4a.

In Figure 8, we present the results of measurements taken from various regions of the same scan. This figure includes the same metrics as in Figure 7 but examines regions of different sizes along the slow-scanning direction (i.e., y-axis): two regions of $3.0 \mu\text{m} \times 0.6 \mu\text{m}$, two regions of $3.0 \mu\text{m} \times 1.5 \mu\text{m}$, and the entire $3.0 \mu\text{m} \times 3.0 \mu\text{m}$ area of the scan. Two observations can be made regarding the accuracy metric shown in Figure 8a. First, there is no clear trend in accuracy across regions of a specific size; rather, the results appear stochastic. Second, accuracy tends to increase with region size, reflecting a greater number of measurements. Conversely, the percentage uncertainty depicted in Figure 8b does not vary with the number of measurements and is determined by the sample's roughness.

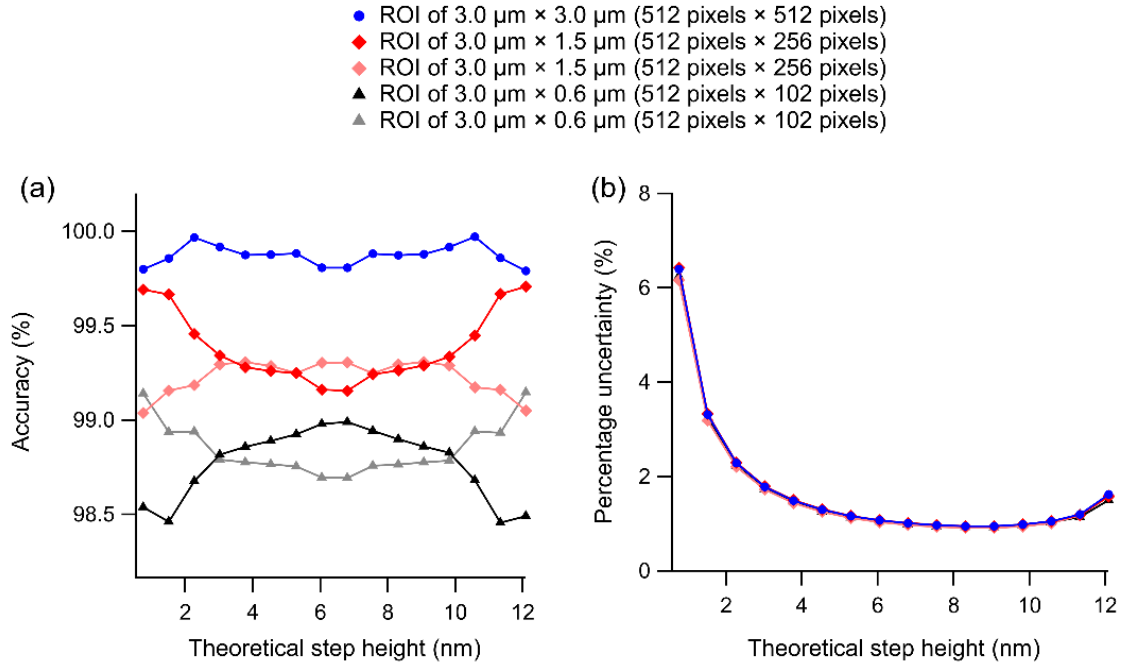


Figure 8. (a) Relative accuracy and (b) percentage uncertainty of the step height calibration applied to regions of interest (ROI) of various sizes on the selected strips shown in Figure 4a.

The overall z-linearity of the calibration within the measured step-height range is evaluated by calculating the slope of the linear fit between the measured heights and their theoretical values. This relationship is illustrated in Figure 9, which presents the combined average values from Table 1 for the trace and retrace measurements. The linear fit was performed using weights that are the inverse of the squares of the standard errors of the measured heights. From the slope of the linear fit, we determine the z-linearity to be $99.99 \pm 0.06\%$, expressed as a percentage relative to one (i.e., the unit slope).

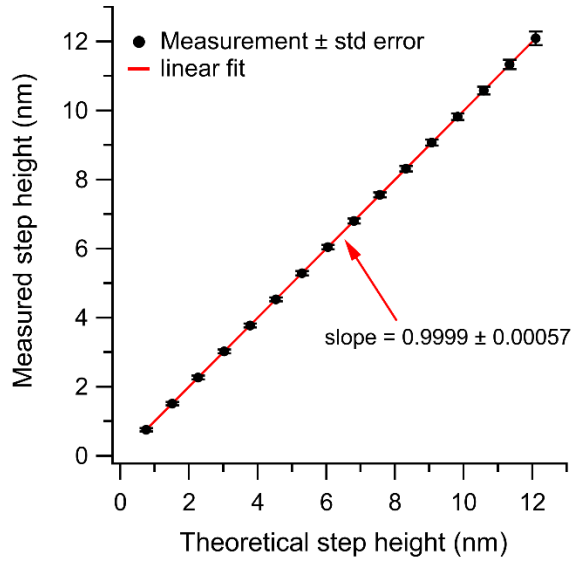


Figure 9. Z-linearity over the calibration range is assessed by a linear fit of the measured step heights to their theoretical values.

We also evaluated the repeatability of the metrics discussed above. Thus, Figure 10 illustrates four key metrics: residual average step height (Figure 10a), standard error (Figure 10b), relative accuracy (Figure 10c), and percentage uncertainty (Figure 10d), all obtained from five consecutive scans over the same area of the 6H-SiC(0001) stepped surface. The scan referenced in the above discussion, covering Figures 2 through 9, is actually the last in this series and is labeled scan #5 in Figure 10. All the data presented in Figure 10 are derived solely from the combined trace and retrace image. For each metric, we observe minimal variation from one scan to another, with no clear correlation between these variations and the scanning order. In fact, the metrics from scan #1 (red markers) and scan #5 (green markers) show the greatest similarity in terms of values and trends with height. In Figure 10a, we can see variations of 1 pm to 3 pm for the smallest step height of 755.76 pm and up to 60 pm for the largest height difference of 12,092.20 pm.

The standard errors associated with these averages show a consistent increase as the step height increases, as illustrated in Figure 10b. The errors range from approximately 50 pm for the smallest height to about 200 pm for the largest height. Across the five scans, the variations in the measured standard errors remain within 5 pm for almost the entire range of step heights. However, these variations become more pronounced for the 12,092.20 pm height difference, likely due to the reduced number of data points used for this height. As mentioned earlier, only two terraces are used for the largest step height: the lowest and the highest in the image.

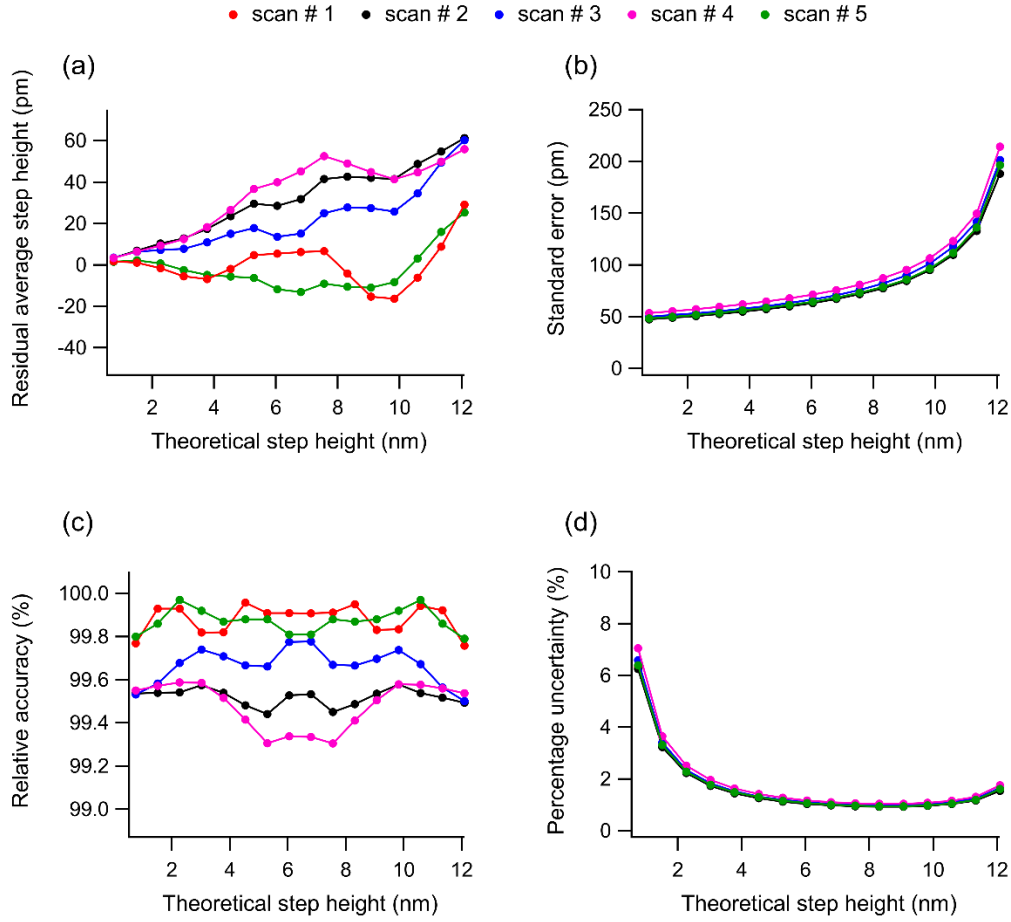


Figure 10. Step height calibration metrics from five consecutive scans: (a) Residual (measured minus actual) average step height differences; (b) The standard error of the measured step height values; (c) The relative accuracy of the measurements; (d) The percentage uncertainty associated with the measurements.

The relative accuracy shown in Figure 10c remains consistent regardless of the scan order. Unlike the average and standard error values, it does not depend on the step height and remains above 99 % across the entire range of step heights. The percentage uncertainty is calculated directly from the standard error, as indicated in equation (6). As shown in Figure 10d, there is a height dependence, but it differs from the standard error depicted in Figure 9b. Although the standard error is smallest for the smallest step height, the percentage uncertainty at this height is actually larger because it is calculated using a smaller measured value. This trend indicates that to investigate even smaller step heights, we will need samples with flat terraces that have

reduced surface roughness. Doing so will help decrease the associated uncertainty of the measured values.

The consistency of the calibration method was verified through measurements taken at different locations on the samples. Figure 11 displays results from three of these measurements. Panels (a), (b), and (c) show the topography images from the retrace portions of the scans, while panels (d), (e), and (f) present the relative accuracy and percentage uncertainty results derived from the combined trace and retrace data. As observed, the topography images are similar across all locations. The relative accuracy shows small variations without any specific trend across the step height range, remaining above 99.5% at each location. Additionally, the percentage uncertainty is consistent across all three cases, as it is mainly determined by the roughness of the sample.

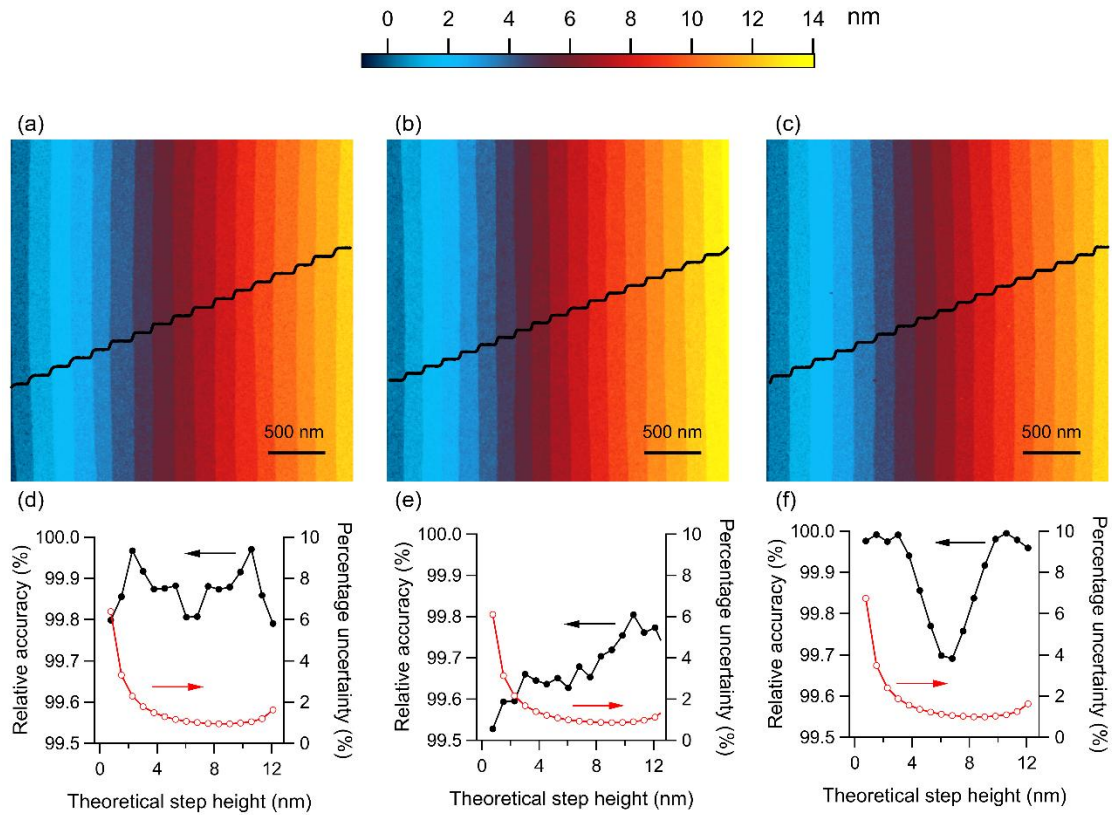


Figure 11. Images of the topography taken from three different locations on a 6H-SiC (0001) sample with step sizes of 755.76 pm are shown in panels (a), (b), and (c). The average line profiles are overlaid on each image. Panels (d), (e), and (f) present the relative accuracy and percentage uncertainty as functions of step height, based on the data collected from the scans shown in (a), (b), and (c), respectively.

4. Discussion

4.1 Potential artifacts and expanded calibration range

In the following, we will examine various topics related to potential artifacts and how we can modify the proposed method to expand the height calibration range. The primary source of artifacts arises from the selection of continuous, flat regions on the surface terraces. Figure 12 illustrates the impact of this selection on the calibration metrics. In Figure 12a, we see the same selection of points (Selection 1) used to define the terraces as in Figure 3b. The metrics plotted in Figures 12c to 12f for Selection 1 are identical to those presented in Figures 6 and 7. In contrast, Selection 2, shown in Figure 12b, incorporates wider portions of the same terraces. Here, the points delimiting these regions (indicated by red markers) are placed closer to the step edges, resulting in strips that are not continuously flat. This leads to significant variations at each terrace level. Consequently, with Selection 2, we observe larger residuals between the measured step heights and their theoretical values, as depicted in Figure 12c. These variations tend to increase with the step height. A notable change is also observed in the relative accuracy calculations when comparing Selection 2 to Selection 1. While the relative accuracy is approximately 99.8 % with Selection 1, it decreases to about 98 % with Selection 2. However, there is almost no difference in the standard errors and percentage uncertainty calculated for Selection 1 and Selection 2. This is because these two metrics are related to surface roughness, which, on average, remains largely unchanged for both selections.

The height range we can achieve with this calibration method is determined by the number of steps within the scanned area. While we could theoretically expand the scan area and enhance spatial resolution to include more steps at a consistent pixel density per terrace, the challenge is that as we increase the number of steps, the terraces are more likely to show misalignment, particularly along the y-axis. This deviation makes it more difficult to use rectangular masks for selecting continuous flat strips along the terraces captured in a scan. As a result, the height range is limited by the number of steps in an image, and flat terraces can still be chosen using simple rectangular shapes across the scanned region. This limitation becomes less significant if the selection method can use irregular shapes that conform to the contours of the terraces [3]. However, for simplicity and to focus on the fundamentals of the method presented here, we will consider only rectangular selective masks.

An example of a scan featuring an increased number of terraces is shown in Figure 13. This scan was acquired at $512 \text{ pixels} \times 512 \text{ pixels}$ over a $10 \text{ }\mu\text{m} \times 10 \text{ }\mu\text{m}$ area of a 6H-SiC(0001) stepped surface with 1511.51 pm steps. As seen in the topography image in Figure 13a, the terraces do not appear straight in the vertical direction across the entire area; instead, they exhibit pronounced curvature at both the bottom and the top of the image. However, within the region outlined by the dashed rectangle, we can retain terraces with straight, parallel edges and analyze them accurately using the calibration method described earlier.

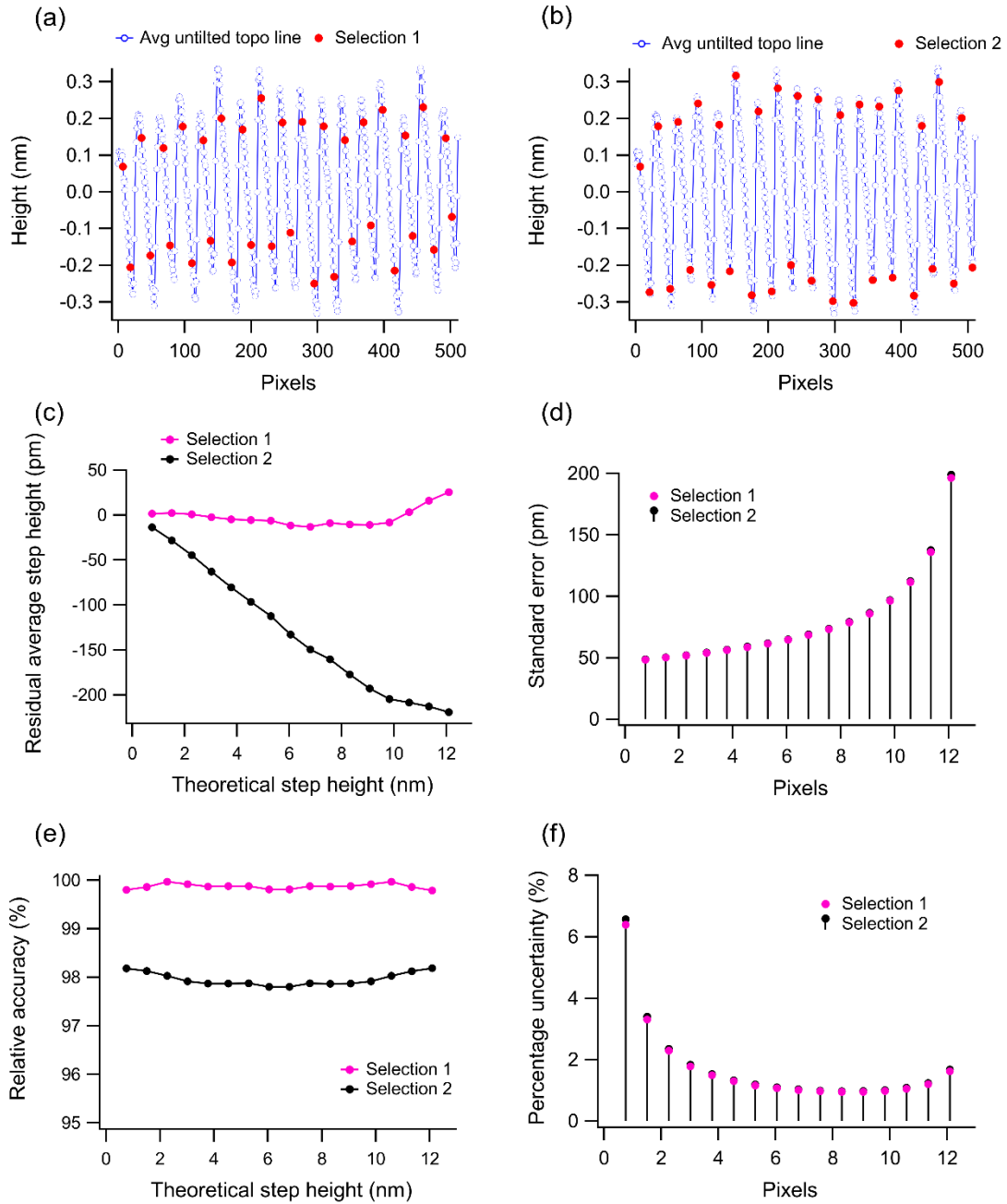


Figure 12. (a) Selection 1 (to get the flat, inner part of the terraces) and (b) Selection 2 (to include points at the end of the terraces) of the terraces from the same average tilted profile. Various metrics were analyzed under the two selections: (c) Residual (measured minus actual) average step height differences; (d) The standard error of the measured step height values; (e) The relative accuracy of the measurements; (f) The percentage uncertainty associated with the measurements.

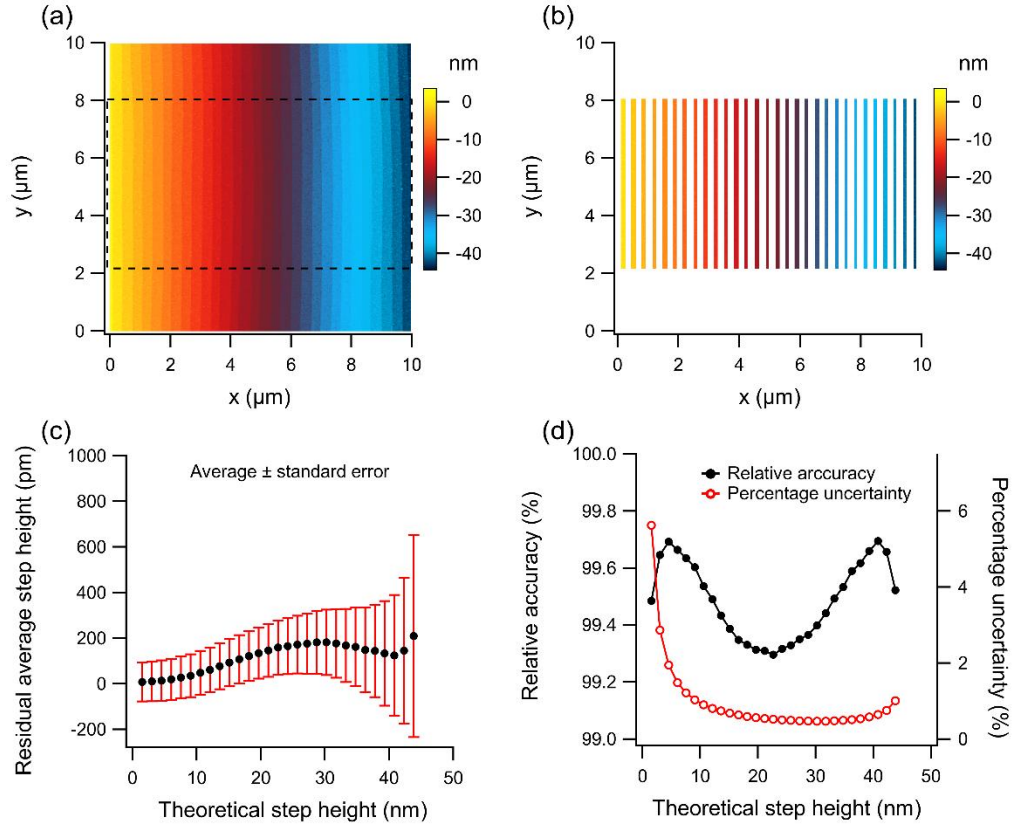


Figure 13. (a) 10 μm x 10 μm area of a 6H-SiC(0001) stepped surface with 1.512 nm steps. (b) Selected flat strips within the selected area shown in (a). (c) Average and standard errors of the step height residuals from data shown in (b). (d) Relative accuracy and percentage uncertainty of the measurements shown in (c).

In both Figures 13a and 13b, the topography and selection of flat strips along the imaged terraces are displayed for the trace image of the scan. However, in Figures 13c and 13d we present the calibration metrics derived from the combined trace and retrace images. As shown in Figure 13b, a total of 30 strips were selected along the terraces within the rectangular region highlighted in Figure 13a. These 30 terraces represent a maximum height difference of approximately 45 nm between the lowest and highest terraces in the image. Figure 13c plots the height residuals determined from the average measured step heights and their associated standard errors against the theoretical values. For step heights up to 30 nm, the average residuals increase gradually from 9 pm to approximately 200 pm, with standard errors of 100 pm. For step heights greater than 30 nm, the average residuals stabilize around 200 pm, while the standard errors can increase up to 500 pm. The overall accuracy and uncertainty of these measurements are illustrated in Figure 13d for all step heights. As shown, the measurements demonstrate relative accuracy exceeding 99 %. For step heights smaller than 10 nm, the

percentage uncertainty in Figure 13d is only a few percent, while it drops below 1 % for step heights greater than 10 nm.

The ability to perform height calibration at multiple points within a specified range provides a valuable assessment of the AFM's z-piezo calibration and linearity across that range. This approach contrasts sharply with calibration methods that focus solely on a single height measurement. Since we require the AFM to track height variations from subnanometers to tens and hundreds of nanometers, multiple-point calibration is essential to ensure both accurate calibration and linearity of the piezo at these points. The step-height calibration method, which uses stepped surfaces, offers a straightforward and effective solution to meet these requirements.

4.2 Detecting miscalibration with the AFM staircase method

The application of our calibration method for detecting z-piezo miscalibration in an AFM is demonstrated in Figure 14. We conducted two successive scans over the same $5.0\ \mu\text{m} \times 2.5\ \mu\text{m}$ area of a stepped 6H-SiC(0001) surface, which features steps measuring 1511.51 pm, which is one full c-lattice constant of 6H-SiC(0001) crystallographic structure [17]. The topography image from one of the scans is shown in Figure 14a. In one scan, we maintained accurate z-piezo calibration, as shown in Calibration 1 of Figure 14. In the other scan, we reduced the z-piezo calibration by 2.5 %, which is labeled Calibration 2 in Figure 14b. In Figure 14b, the average topography profile across the imaged steps is displayed for these two different calibrations. It is clear that the difference between these two calibrations, and consequently the step height value, increases as we move away from an imposed matching point, which is located at the left side of the plot. We cannot determine which calibration method is correct solely from the average profiles. However, by examining the residual values (the difference between measured and actual step heights) shown in Figures 14c and 14d, it becomes evident that Calibration 1 provides consistent results across the entire height range. In contrast, Calibration 2 shows increasing miscalibration with increasing height. The impact of this miscalibration is also reflected in the relative accuracies of the two calibrations, which are illustrated in Figures 14e and 14f. Calibration 1 demonstrates a relative accuracy of over 99.5 %, while Calibration 2's relative accuracy falls to 96 %. Indeed, the calculated z-linearity from the slope of measured versus theoretical step heights across the entire calibration range was $(99.54 \pm 0.08)\%$ for Calibration 1 and $(95.60 \pm 0.09)\%$ for Calibration 2. It is noteworthy that the relative accuracy remains within 1% across the entire height range, even when the z-piezo calibration is modified. Although the piezo's response may be miscalibrated, it still demonstrates linear behavior. This suggests that the accuracy's dependence on height reflects the linearity of the z-piezo within that range. In contrast, a linear or parabolic relationship between accuracy and height would indicate a non-linear response of the piezo with height. Additionally, it is important to note that both the standard errors (represented by error bars in

Figures 14c and 14d) and the percentage uncertainties (shown in Figures 14e and 14f) remain unaffected by this miscalibration; they yield the same values for both Calibration 1 and Calibration 2. As before, all the calibration metrics were determined by analyzing the combined trace and retrace images of the scans.

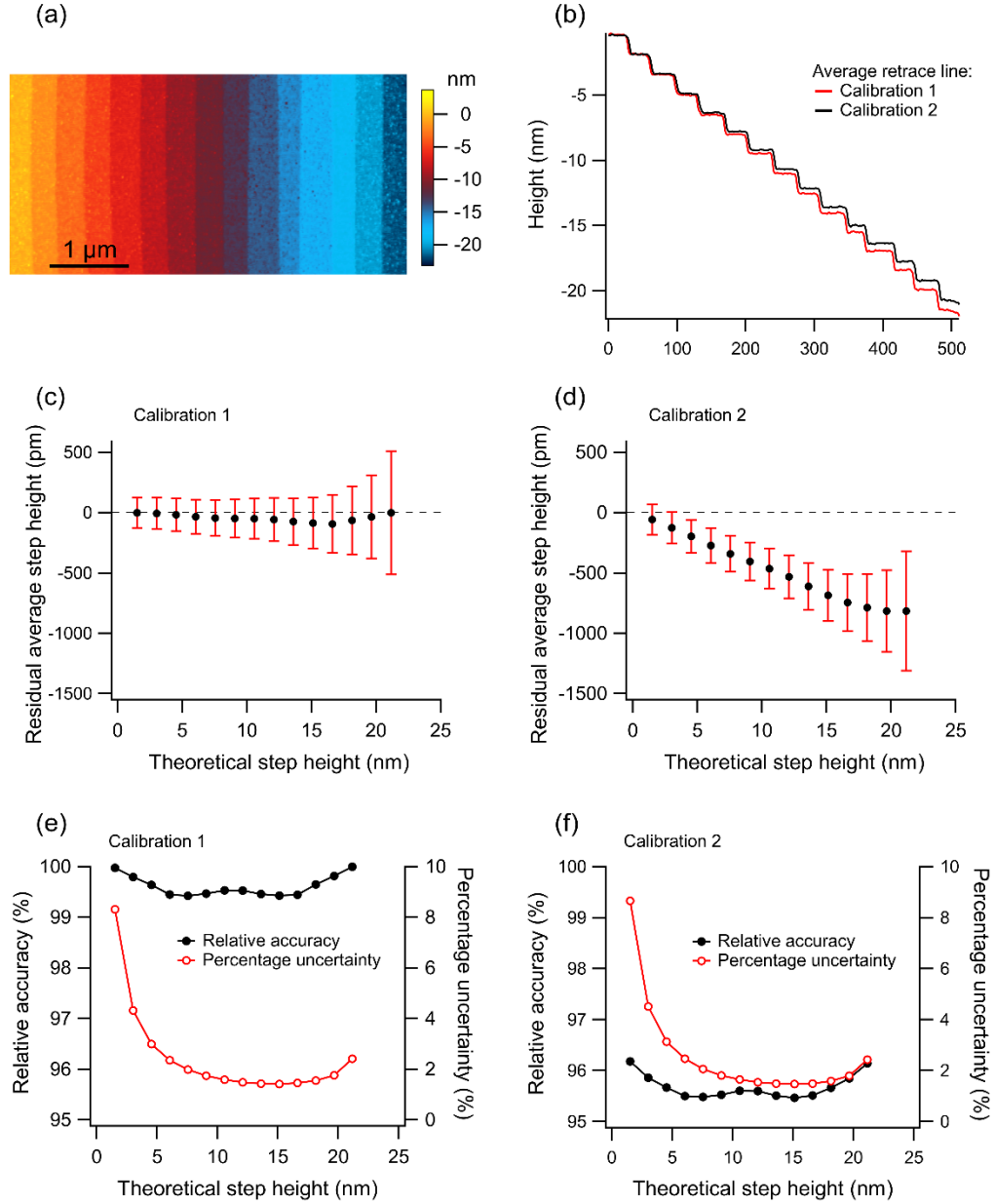


Figure 14. (a) 5.0 $\mu\text{m} \times 2.5 \mu\text{m}$ area of a 6H-SiC(0001) stepped surface with 1.512 nm steps. (b) Average retrace line with a 2.5 % difference between z-piezo calibrations. (c) and (d) Residual step height differences (measured minus actual) for Calibration 1 and Calibration 2, respectively. (e) and (f) Relative accuracy and percentage uncertainty for Calibration 1 and Calibration 2, respectively.

The analysis presented above shows that, among the calibration metrics, the residual values and relative accuracy are the most sensitive indicators for detecting miscalibration of the z-piezo across various step heights. Low relative accuracy values indicate poor calibration, while residual values quantify the extent of miscalibration at each step height. These results clearly demonstrate the advantages of multi-point calibration over single-point calibration.

The 6H-SiC(0001) surfaces with step heights of 755.76 pm were obtained from Ted Pella (Redding, CA, USA), while those with step heights of 1511.51 pm were sourced from K-TEK (Wilsonville, OR, USA). Measurements on the surfaces with 755.76 pm steps (data, figures, and discussions detailed in Sections 2, 3, and 4.1) were conducted using a Vero AFM (Oxford Instruments/Asylum Research, Santa Barbara, CA, USA). Measurements on the surfaces with 1511.51 pm steps (discussed in Sections 4.1 and 4.2) were made with a MultiMode 8 AFM (Bruker, Santa Barbara, CA, USA). The AFM probes used in this study were Olympus AC160TS-R3 (Oxford Instruments/Asylum Research, Santa Barbara, CA, USA) on Vero AFM and PPP-SEIH (Nanosensors, Neuchatel, Switzerland) on MultiMode 8 AFM. All the scans were performed in AFM amplitude modulation (tapping) mode.

5. Conclusion

In this study, we present a straightforward calibration method for the z-piezo transducer of an AFM used to measure heights in the sub-nanometer to tens-of-nanometer range. The method utilizes the uniform distribution of step-terrace arrays found on stepped surfaces, such as 6H-SiC(1000). By incrementally adding the height of steps captured in an AFM scan, we can verify multiple checkpoints across the specified range. The metrics used in this method are based on residual differences between measured and theoretical values, relative accuracy, and percentage uncertainty. These metrics can assess both the vertical calibration and linearity of the z-piezo within the specified range. Although this method is illustrated using SiC steps, it can be applied to any similar stepped surfaces, allowing even smaller lattice distances to be used as unit heights for calibration. Utilizing the natural distances between crystallographic planes as metrological references for height measurements is particularly significant at the subnanometer scale, where dimensional artifacts are not available.

Data availability statement

The source code for the calibration procedure described in this work can be found in the *AFM-Staircase* NIST GitHub repository, <https://github.com/usnistgov/AFM-Staircase> [18].

Author contributions

Gheorghe Stan – orcid.org/0000-0001-5773-0328;
Email: gheorghe.stan@nist.gov

Disclaimer

Certain commercial equipment, software, instruments, and materials are identified in this document. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

Acknowledgments

This work was performed with funding from the CHIPS Metrology Program, part of CHIPS for America, National Institute of Standards and Technology, U.S. Department of Commerce.

References

1. A. Yacoot, H. Bosse, and R. Dixon, The lattice parameter of silicon: a secondary realisation of the metre, *Meas. Sci. Technol.* 31, 121001 (2020).
2. J. Garnes, D. Necas, L. Nielsen, M. H. Madsen, A. Torras-Rosell, G. Zeng, P. Klapetek, and A. Yacoot, Algorithms for using silicon steps for scanning probe microscope evaluation, *Metrologia* 57, 064002 (2020)
3. M. A. Lawn, Z. Bolton, L. Murphy, S. Gartner, Y. Oh, and V. A. Coleman, Step height measurement of monoatomic silicon crystal lattice steps with a commercial atomic force microscope, *Meas. Sci. Technol.* 35, 105018 (2024).
4. ISO 11952:2019, Determination of geometric quantities using SPM: Calibration of measuring systems, ed. 2 (2019).
5. C. Wang, L. Li, W. Jing, Y. Zhang, S. Wang, Q. Lin, D. Xian, Q. Mao, Y. Zhang, D. Duan, M. Liu, and Z. Jiang, A Study on the Sub-5 nm Nano-Step Height Reference Materials Fabricated by Atomic Layer Deposition Combined with Wet Etching, *Micromachines* 13, 1454 (2022).
6. T. Ohmi; S. Aoyama, Calibration of height in atomic force microscope images with subnanometer scale silicon dioxide steps, *Appl. Phys. Lett.* 61, 2479–2481 (1992).
7. M. Suzuki, S. Aoyama, T. Futatsuki, A. J. Kelly, T. Osada, A. Nakano, Y. Sakakibara, Y. Suzuki, H. Takami, T. Takenobu, M. Yasutake, Standardized procedure for calibrating

- height scales in atomic force microscopy on the order of 1 nm, *J. Vac. Sci. Technol. A* 14, 1228–1232 (1996).
8. J. Fu, V. Tsai, R. Koning, R. Dixon, T. Vorburger, Algorithms for calculating single-atom step heights, *Nanotechnology* 10, 428 (1999).
 9. H. Matsunami and T. Kimoto, Step-controlled epitaxial growth of SiC: High quality homoepitaxy, *Mater. Sci. Eng. R.* 20, 125–166 (1997).
 10. S. Nie, C. D. Lee, R. M. Feenstra, Y. Ke, R. P. Devaty, W. J. Choyke, C.K. Inoki, T. S. Kuan, and G. Gu, Step formation on hydrogen-etched 6H-SiC{0001} surfaces, *Surface Science* 602, 2936–2942 (2008).
 11. T. V. Vorburger, A. Hilton, R. G. Dixon, N. G. Orji, J. A. Powell, A. J. Trunek, P. G. Neudeck, and P. B. Abel, Calibration of 1-nm SiC step height standards, *Proc. SPIE* 7638, Metrology, Inspection, and Process Control for Microlithography XXIV, 76381D (2010).
 12. S. Smink, L. N. Majer, H. Boschker, J. Mannhart, and W. Braun, Long-Range Atomic Order on Double-Stepped Al₂O₃(0001) Surfaces, *Adv. Mater.* 36, 2312899 (2024).
 13. V. W. Tsai, T. Vorburger, R. Dixon, J. Fu, R. Köning, R. Silver, and E. D. Williams, The study of silicon stepped surfaces as atomic force microscope calibration standards with a calibrated AFM at NIST, *AIP Conf. Proc.* 449, 839–842 (1998).
 14. P. J. Mohr, B. N. Taylor, and D. B. Newell, CODATA recommended values of the fundamental physical constants: 2022 Rev. *Mod. Phys.* 97, 025002 (2025).
 15. A. Yacoot, L. Koenders, and H. Wolff, An atomic force microscope for the study of the effects of tip–sample interactions on dimensional metrology, *Meas. Sci. Technol.* 18, 350–359 (2007).
 16. ISO 5436-1:2000 Geometrical Product Specifications (GPS)—Surface Texture: Profile Method; Measurement Standards—Part 1: Material Measures (2000).
 17. T. Kimoto, and J. A. Cooper, Appendix C Major physical properties of common SiC polytypes, *Fundamentals of Silicon Carbide Technology: Growth, Characterization, Devices and Applications* (Wiley) pp 521–4 (2014).
 18. G. Stan, AFM-Staircase, <https://github.com/usnistgov/AFM-Staircase>, (2026).
 19. G. Stan, AFM Staircase Calibration, National Institute of Standards and Technology, <https://doi.org/10.18434/mds2-4056>, (2026).