

AI-equipped scanning probe microscopy for autonomous site-specific atomic-level characterization at room temperature

Zhuo Diao,^{1, a)} Keiichi Ueda,² Linfeng Hou,¹ Fengxuan Li,¹ Hayato Yamashita,¹ and Masayuki Abe^{1, b)}

¹⁾ Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-0043, Japan

²⁾ Tokyo Metropolitan Industrial Technology, Research Institute, 2-4-10 Aomi, Koto-Ku, Tokyo, 135-0064, Japan

We present an advanced scanning probe microscopy system enhanced with artificial intelligence (AI-SPM) designed for self-driving atomic-scale measurements. This system expertly identifies and manipulates atomic positions with high precision, autonomously performing tasks such as spectroscopic data acquisition and atomic adjustment. An outstanding feature of AI-SPM is its ability to detect and adapt to surface defects, targeting or avoiding them as necessary. It's also engineered to address typical challenges such as positional drift and tip apex atomic variations due to the thermal effect, ensuring accurate, site-specific surface analyses. Our tests under the demanding conditions of room temperature have demonstrated the robustness of the system, successfully navigating thermal drift and tip fluctuations. During these tests on the Si(111)-(7×7) surface, AI-SPM autonomously identified defect-free regions and performed a large number of current-voltage spectroscopy measurements at different adatom sites, while autonomously compensating for thermal drift and monitoring probe health. These experiments produce extensive data sets that are critical for reliable materials characterization and demonstrate the potential of AI-SPM to significantly improve data acquisition. The integration of AI into SPM technologies represents a step toward more effective, precise and reliable atomic-level surface analysis, revolutionizing materials characterization methods.

Keywords: self-driving, deep learning, scanning probe microscopy, scanning tunneling spectroscopy, room temperature

INTRODUCTION

The integration of artificial intelligence (AI) with nanotechnology has been recognized as crucial since the pioneering work by Drexler in 1986¹. However, the advancements described in this book had not been widely realized. As a key tool of the nanotechnology, scanning probe microscopy (SPM) has emerged in characterizing nanoscale surfaces, enabling the discovery of new surface properties and phenomena^{2,3}. Today, SPM has significantly contributed to both basic science and industrial applications. Especially, in the field of basic science, it can not only image surfaces, but also measure the physical properties of individual atoms and move atoms to create structures⁴. Most of these experiments have been conducted in cryogenic environments. This is because not only the SPM equipment but also the tip of the probe and the sample itself are thermally stable. However, from a practical standpoint, conducting these processes at room temperature is essential.

Even in room temperature environments, where thermal effects can affect measurements, SPM has provided significant capabilities such as dynamic imaging to observe temporal changes in chemical reactions^{5,6}, biological processes⁷, surface dynamics^{8,9}, diffusion^{10,11}, and crystal growth¹². In addition, there are studies of dopant atom manipulation that can be done at room

temperature^{13,14}.

There have been challenges in further pursuing these pioneering room-temperature experiments. Using the room-temperature SPM to achieve atomic resolution does not remove the thermal effects on the device, which lead to measurement instability. One of the most significant effects of thermal fluctuations in atomic-resolution SPM measurements is thermal drift. Thermal drift continuously changes the relative position of the tip and sample atoms, which not only prevents continuous measurement of the same area, but can also cause image distortion. Another problem of the room-temperature SPM is frequent change of the tip apex. The quality of the image is affected by the frequently change of the tip apex atom especially in atomic resolution imaging. Repairing the tip apex is usually performed by touching to the surface, which requires time and a great deal of attention. To achieve the high-precision measurements under such inherently non-optimized conditions, SPM techniques such as drift correction^{15,16} and tip fabrication^{17,18} become indispensable in the context of room temperature SPM. However, even with these techniques, it was very difficult to perform site-specific experiments at the atomic level at room temperature.

In recent years, to overcome the limits of human capability, experiments utilizing the concept of the self-driving laboratory^{19,20} have become increasingly prevalent, driven by the demand for extensive and complex data collection. This approach has risen to prominence as a key solution for challenges like new materials discovery, largely through the application of artificial intelligence (AI). The self-driving concept is also anticipated to be

^{a)}enzian0515@gmail.com

^{b)}abe.masayuki.es@osaka-u.ac.jp

highly beneficial for time-consuming and labor-intensive SPM experiments, which is manually operated. A deep learning model, adept in signal processing and computer vision, can identify specific patterns with accuracy nearly on par with human experts^{21,22}. Strategically integrating AI can significantly reduce the dependency on manual intervention in SPM operations²³.

There are previous studies of the integrating the image recognition AI and SPM yielding notable contributions to the domains of data analysis and processing^{24–28}. The application of these AI-driven analysis techniques in the processing of data measured in real-time holds the potential to significantly automate SPM operations. The efficacious deployment of such technologies requires intricate integration within the control mechanisms, encompassing both hardware and software, that includes scanning protocols and AI features, alongside the acquisition of extensive datasets for the cultivation of high-efficiency AI models. Particularly in cryogenic conditions, where thermal disturbances are markedly reduced, recent research has underscored the utility of AI-facilitated methods in promoting autonomous scanning procedures^{29–31} and the manipulation of individual atoms³². However, unlike setups in cryogenic environments, more sophisticated AI-based SPM (AI-SPM) is required in room temperature environments. This enhancement is necessary to compensate for challenges posed by thermal fluctuations and to ensure the reproducibility of experiments. Robustness in handling measurements under unstable conditions is crucial for performing real-time and site-specific experiments at room temperature. Moreover, it is essential to recognize during dataset collection that the simulated datasets used for training^{25,27} may not fully capture the variations in atomic images. This limitation arises from differences in the tip state and the patterns of adsorbates in actual measurements. Addressing these considerations is vital for enhancing the reliability and performance of site-specific SPM measurements at room temperature.

In this manuscript, we present a deep learning-assisted AI-SPM system, specifically designed for the site-specific operation at room temperature. Our AI-SPM system is an integrated fusion of software, control firmware, and hardware components, which facilitates the development of a robust neural network through a systematic dataset collection process. Decision-making and optimization tailored for room temperature conditions, is fully automated, with the objective of achieving atomic precision measurements. Utilizing this system, we demonstrate two key applications in surface characterization: autonomous acquisition of high quality images and large data sets for atomic precision scanning tunneling spectroscopy (STS) at room temperature.

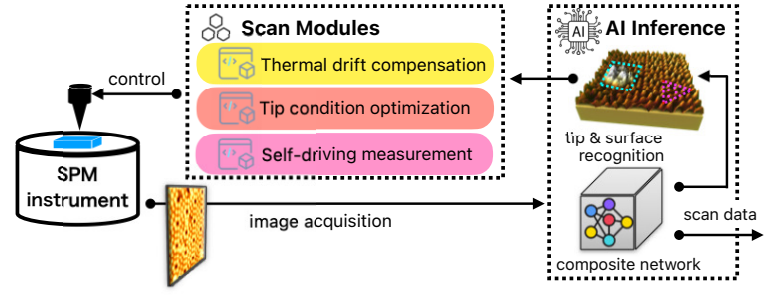


FIG. 1. Schematic flow of artificial intelligence scanning probe microscopy (AI-SPM) system. The system compensates for thermal drift and the tip apex, which is a problem inherent to room temperature. Then, It also determines the presence of adsorption sites and defects, and performs site-specific experiments at targeted atomic positions. Large amounts of data can be obtained automatically.

AI-SPM FOR SELF-DRIVING MEASUREMENT

AI-SPM configuration

Figure 1 illustrates the configuration of our AI-SPM system. SPM hardware is conventional one, and two main program parts are added to our home-built scan software for the AI operation. The AI inference section receives the SPM measurement data, makes a situational judgment and determines the next task. In particular, it uses the images to determine the state of the probe tip, the identification of individual atomic sites and unit cells, the location and type of adsorbates, and whether further site-specific measurements are possible. The information of the tip and sample surface sent to the Scan Module part from the AI inference component is utilized in the experiment programmed by the operator. In Fig. 1, the Scan Module comprises two scripts essential for the room temperature experiment (thermal drift compensation and tip condition optimization) alongside a self-driving measurement script. These scripts are used in the experiments in this manuscript.

Convolutional neural networks on AI inference

The acquired image data of the surface not only confirms the crystal structure of the surface, but also provides information of the condition for the site-specific measurement: presence of defects or adsorbates, not atomically clean area or steps, tip apex condition. To automatically make decisions on all this information, we have employed convolutional neural networks (CNNs). Each CNN is tailored for a specific predictive task, and they all utilize scanned topography as their input. By integrating these CNNs into a composite network, we enable comprehensive access to a wide range of information about the real-time scanning topography³³. Understanding of both tip and surface conditions enables

the site-specific measurements such as STS and atomic manipulation.

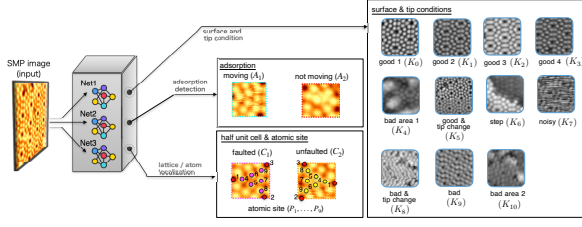


FIG. 2. Convolutional neural networks (CNNs Net1, Net2, and Net3) employed in this study for room-temperature site-specific analysis of the Si(111)-(7×7) surface. Net1 is tasked with recognizing the conditions of the tip and sample, Net2 identifies adsorption sites, and Net3 is responsible for recognizing half unit cells and atomic positions.

In Fig. 2, the CNN architecture for the Si(111)-(7×7) surface is presented. We designed a composite network structure comprising three distinct models—Net1, Net2, and Net3—each dedicated to a specific task. Net1 is tasked with recognizing the conditions of both the tip and sample, performing a multi-class classification of tip and surface conditions $K_i (i = 0, 1, \dots, 10)$ ³³. It determines whether the sample and the probe are both in a state ($K = 1, 2, 3, 4$) in which site-specific measurement is possible. If the surface is not contaminated and the tip is capable of atomic resolution, it is considered to be in a “good” condition for site-specific measurement. Our previous studies have demonstrated that the accuracy of Net1 is from 87% to 90%. The reason for classifying good into four categories is to determine and create a probe that can be used for atomic manipulation in the future^{34,35}.

Net2 determines adsorbates. On the Si(111)-(7×7), we can find some adsorbates at the atomic level that are moving and some are not. Here, we define the one imaged brighter than the Si adatom in the half-unit cell to be the immobile adsorbent A_2 . As another type of adsorbate, there are ones that are imaged like noise or feedback error. These are adsorbates in motion remaining in a half-unit cell (A_1 in Fig. 2). The moving adsorbates in the half unit cell have been studied for experiments with clusters and atomic manipulation^{36–44}. Net2 is designed to be applied to these studies as well.

Net3 detects site-specific information of the Si(111)-(7×7) surface. Half-unit cells are categorized as bounding boxes of C_1 and C_2 , respectively. Applying a negative sample bias during measurement and observing the screen’s response allows for the determination of whether C_1 or C_2 corresponds to the faulty or non-faulty section. In the experiment of Fig. 2, C_1 was identified as faulted, while C_2 was found to be unfaulted. Local atomic sites are represented as key points $P_i (i = 1, 2, \dots, 9)$ (points of coordinate (x, y)): three corner holes around each half-unit cell as P_1, P_2, P_3 , corner adatom as P_4, P_8, P_9 , and center adatom as P_5, P_6, P_7 .

Scan Module Scripts

The Scan Module comprises two scripts essential for the room temperature experiment (thermal drift compensation and tip condition optimization) alongside a self-driving measurement script. These scripts are used in the experiments in this manuscript. Users can add own script to the Scan Module to customize the experiment.

The module for the thermal drift compensation works based on the feedforward technique present in our previous study⁴⁵. Continuously obtained SPM images are compared using the feature point matching algorithm to output the thermal drift velocity at the minute scale, enabling the correction of even non-linear thermal drift at the days scale.

It is equipped to autonomously maintain the optimum state of the probe³³. Here, the CNN determines the state of the tip from the images obtained: if the CNN determines that the tip is not optimal for atomic resolution measurements, a tip shaping is performed by bringing the tip close to the surface and simultaneously changing the bias current. The images are then acquired and the Net1 judges the state of the tip. These processes can be performed automatically until the tip becomes a “good” condition. The thermal drift module is activated during the tip optimization process for automatically compensating for thermal drift.

The self-driving measurement module autonomously performs data acquisition based on the AI Inference output and can also leverage other module functionalities for drift correction and tip condition optimization. Therefore, our AI-SPM system represents a departure from traditional automation methods with fixed routines. It offers the ability for self-guided data exploration and acquisition, adapting to previous scans through a closed-loop mechanism. Detailed descriptions of the AI-SPM hardware are provided in the Methods section.

Two-phase training data acquisition

As mentioned above, at room temperature, even if thermal drift is compensated, we are facing unstable image conditions due to tip apex change. Adsorbates and defects may be present in the first place, but they may also appear during scanning, significantly influencing the SPM images. This makes us assemble a comprehensive and varied set of training data is needed. To acquire data that take into account both the number and variety of datasets, we have employed developed “two-pase” approach to training the AI-SPM itself as the experiment progresses.

Phase 1 comprises collecting and creating a dataset comprising image information obtained under diverse conditions. Here, every image is acquired, including those that have not been atomically resolved. To obtain a lot of images categorized K_0 to K_{10} , same process used in the tip-condition optimization was conducted. In

this phase, although the initial accuracy may not fully optimize the tip and sample state, this process continues iteratively, accumulating an extensive dataset. This dataset is essential in training a robust model that can distinguish data quality, ensuring the system's capability to autonomously capture atomic-resolution images on Si(111)-(7×7).

The second phase is intended to further enhance the data set of images being obtained at atomic resolution. Autonomous measurement on a target atom is performed within the self-driving measurement module, with the assistance of Net3, which provides information about atom locations. One of the examples of measurement is autonomous atom manipulation techniques. It utilizes the "atomic pen" technique¹⁴ to manipulate Si atoms within a unit cell and individually place in-motion Si atoms as single adsorptions. By repeatedly executing phase 2, a larger number of adsorbate and defect patterns on the topography are generated, augmenting the datasets utilized in training Net1, Net2, and Net3.

To enhance the accuracy of Net1 and Net3, the dataset has been expanded by incorporating topography scans obtained from various leading states and scan areas. In phase 1 and phase 2, 11,616 and 7,269 images were acquired, respectively. From these, 2,082, 255, and 545 images and their augmented data were used in the training of Net1, Net2, and Net3, respectively. These data points, likely consisting of individual images or instances, comprehensively cover a wide range of surface properties. The performance evaluation using confusion matrices of Net1, Net2, and Net3 is detailed in the supplementary information (see Fig. 6 in Supplementary I), demonstrating that they achieved high scores.

IMPLEMENTATION OF AI-SPM

Local site identification

Performance of the local site identification of STM measurement at room temperature on Si(111) – (7 × 7) surfaces is shown in Fig. 3. Figure 3 shows the capability of the trained Net1, Net2, and Net3 that recognize key points of the surface for the site-specific measurements. Images acquired as shown in Fig. 3 (a) are evaluated by Net1. In this image, the (7 × 7) structure and defects and adsorbates are present. For this image, Net1 outputs weight values of $k_2(\text{Good2}) = 0.53$ and $k_4(\text{Bad area}) = 0.47$. This means that the tip condition is good, but surface defects and adsorbates are present. In identifying adsorbates in Net2, it can be seen in Fig. 3 (b) that it can identify stationary adsorbates (A_1), surrounded by red dashed lines, and adsorbates diffusing in the half-unit cell (A_2), surrounded by light blue dashed lines. In Fig. 3 (c), Net3 classifies almost all regions of the image into C_1 (faulted half) and C_2 (unfaulted half) categories, as well as identifying individual adatoms and corner holes within these halves. Net3 is

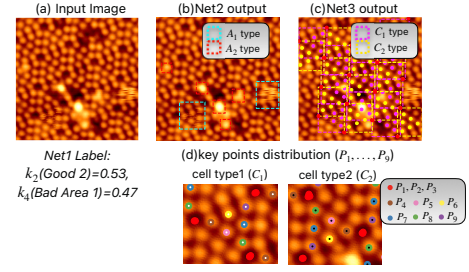


FIG. 3. Performance of the local site identification on Si(111) – (7 × 7) surfaces. (a) an STM image after Net1 is evaluated, which is used for the site identification as shown in (b) to (d). The evaluation results are shown below the image. (b) Identification of adsorbates by Net2. A_1 and A_2 indicate moving and stationary adsorbates, respectively. (c) Net3 classification of half-unit cell type. C_1 and C_2 are faulted and unfaulted half unit cells, respectively. (d) Site-specific identification of adatom site and corner holes in the half-unit cells. Each site is labeled by Net3 as $P_i (i = 1 \cdots 9)$. Sample bias voltage and tunneling current in the STM measurement are $V_s=1.5$ V and $I_t=200$ pA, respectively.

capable of identifying each adatom. As depicted in Fig. 3 (d), the individual adatoms and corner holes are represented by $P_i (i = 1, 2, \cdots, 9)$. The labels indicate that $i = 1, 2, 3$ correspond to corner holes, $i = 4, 8, 9$ to corner adatoms, and $i = 5, 6, 7$ to center adatoms. These results mean that our proposed Net1, Net2, and Net3 methods can identify individual atomic sites as keypoints at room temperature.

Probing for optimal measurement regions

Figure 4 shows a continuous image of the room temperature STM measurement on the Si(111) – (7 × 7) surface, in which atomic images free of defects, adsorbates, and steps are autonomously identified. In the experiment, a total of 45 consecutive images were acquired over different regions. In Fig. 4(a), scanning started from the upper left corner of the figure, with the upper left portion of each image marked by an inverted triangle. The trajectory of the regions scanned is shown by red lines connecting the sequence of images. Here, the scanning routine includes two critical modules to identify the atomic resolution images. First, a thermal drift correction module compensates thermal drift to minimize the image distortion⁴⁵. As the second module, Net1 keeps the tip in a state conducive to atomic resolution measurements while bypassing areas affected by impurities, atomic defects, and step edges. Figure 4(b) shows Net1 judgment of the acquired STM images, classifying the most plausible probe or surface conditions shown in Fig. 2 by color.

Until now, it has been necessary to determine the appropriate measurement area and to continuously monitor the state of the probe tip. However, unlike in cryo-

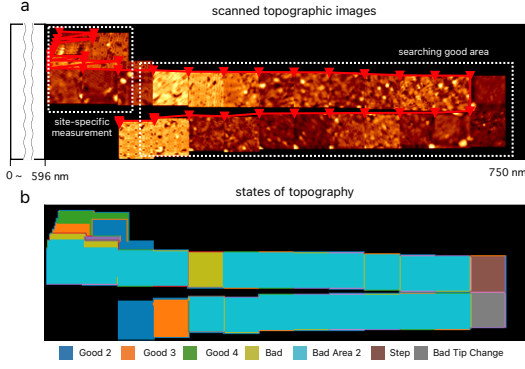


FIG. 4. Sequential images of STM on the Si(111) – (7×7) surface being measured while automatically searching for the optimum measurement area and tip conditions. (a) Sequential images measured by self-guided STM. The first scan starts at the upper right corner, and the scan area is automatically changed. The upper left portion of each image is indicated by an inverted triangle, and the inverted triangles are connected by red lines to show the trajectory. (b) Top-most inferred state during the whole data acquisition process including tip conditioning, drift compensation, and STS target atom localization during continuous scan expressed in color. All STM topographic images were acquired with 2 V sample bias, -200 pA set point, 105 s scan time, and 11.25×11.25 nm scan area. During this measurement, thermal drift correction is activated by feature point matching⁴⁵.

genic environments where the probe tip remains relatively stable, at room temperature the probe tip frequently changes. This makes it virtually impossible for researchers to conduct sustained long-time experiments. As demonstrated in Fig. 4, by integrating advanced automation and deep learning techniques, we have been able to establish optimal experimental conditions for exploring regions with atomic resolution. This method enables the automatic identification of the appropriate probe and region at room temperature, which is essential for conducting site-specific experiments. Consequently, experiments such as atomic manipulation and STS are expected to be feasible under conditions comparable to those in cryogenic environments.

Self-driving scanning tunneling spectroscopy at room temperature on different Si(111) adatom sites

Using the methods described so far, the basic tools are now available to realize an AI-SPM capable of self-steering measurements. As an example of a site-specific measurement at room temperature, we perform $I - V$ measurements on adatoms of the Si(111)-(7×7). From the obtained $I - V$ curve, the data of dI/dV , or STS as the local electronic state, can be calculated. There have been previous studies of the density state of Si(111)-(7×7) surfaces both at room⁴⁶⁻⁴⁹ and low^{49,50} temperatures. Previous studies have reported that the Si surface has different electronic state behaviors depending on tempera-

tures.

The results suggest the importance of performing the site-specific measurement at room temperature. On the other hand, in the room temperature environment, in addition to the thermal drift and tip instability described above, there is also the effect of thermal fluctuations in the LDOS itself due to the broadening of the Fermi function for the population of electrons, resulting in a reduction in the energy resolution of the STS measurement. Therefore, in order to obtain a reliable site-specific measurement, it is necessary to acquire a large amount of data and statistically process the acquired data to deal with irregularities such as data variations and probe changes specific to room temperature. For site-specific measurements at room temperature, the acquisition of large amounts of data has been very difficult with previous room temperature STM setups.

To ensure the acquisition of reliable data in this inherently uncertain environment, we've applied our AI-SPM to perform a large number of $I - V$ curve measurements on the four different adatom sites of the Si(111)-(7×7) and performed statistical analysis to calculate the STS data. Specifically, it can locate individual atomic sites of the respective center and corner adatoms of each faulted and unfaulted unit cell where adsorptions and defects were absent.

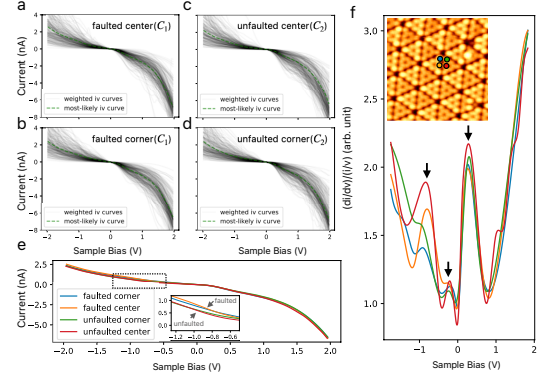


FIG. 5. $I - V$ and calculated STS curves on different adatom site of the Si(111)-(7×7) measured at room temperature. Weighted plot of $I - V$ curves on (a) faulted center adatom (b) faulted corner adatom (c) unfaulted center adatom (d) unfaulted corner adatom. The green dashed lines are the most likely ones calculated by the overall data-based selection method. (e) Over-wrapped $I - V$ curves of the four different adatom sites that are the same as the green dashed curves in (a) to (d). (f) STS (dI/dV) curves calculated from the four $I - V$ curves in (e). The inset in (f) is STM topographic image acquired with the -1.4 V sample bias, 200 pA tunneling current and the scan range is 11.25 nm^2 . The four points on the topography represent the $I - V$ measurement position. The color in the topography points and dI/dV line correspond to each other. the measurement noise is reduced by the Savitzky-Golay and mean filters.

Using this approach, we acquired a total of 324 $I - V$ curves measured at center and corner adatoms within

both faulted and unfaulted unit cells, respectively. Following these measurements, the tip apex condition is evaluated to ensure that it remains optimal, and then the thermal drift is corrected again. $I - V$ data were automatically obtained in appropriate regions and tip conditions as evaluated by AI-SPM. This entire procedure is performed iteratively, enabling the generation of robust and reliable data under the challenges of room temperature.

Obtained $I - V$ curves are individually plotted in Fig. 5(a), (b), (c), and (d). Our preliminary experiments (see Appendix B) have shown that the state of the tip apex changes about 7% of the time when the voltage is swept. To determine the most representative values for the $I - V$ curves of individual atoms (Fig. 5(e)), we employed the data selection approach to obtain a group of curves with the highest tendencies, and the mean $I - V$ curves of the extracted group (see Methods). dI/dV curves were calculated by averaging the extracted $I - V$ curves followed by the mean filter, Savitzky-Golay⁵¹ filter, and smoothing by the 1D spline function to minimize the noise. When zooming in on the region with negative sample bias, subtle differences between the curves of faulted unit cells and unfaulted unit cells become apparent. In this region, the conductance of the faulted unit cells is noted to be greater than that of the unfaulted unit cells, supporting the higher electron density on the faulted side with a stacking fault than on the unfaulted side⁵².

As shown in the inset of Fig. 5(f), in the STM measurements of the Si(111) surface, the contrast between the corner adatom and the center adatom in the faulted and unfaulted half-unit cells is different in the negative sample bias region. The calculated STS results also show that, under the negative sample bias region, the four site-specific locations at faulted/unfaulted corner and center adatom show distinct variations (the color of the points representing the positions of the four atoms corresponds to the color of the dI/dV curves). Because of the charge transfer from adatom to rest atom, the center adatom, which has a larger number of neighbors to the rest adatom, has a lower electron density than the corner adatom⁴⁸. This trend of the four sites is revealed in our STS curves, and the positions of peaks [downward arrow in Fig. 5(f)] align with the past result⁴⁹. Additionally, the most prominent peaks on the surface states of Si(111)-(7 × 7) appear at -0.8 V and 0.3 V, supported by the ultraviolet photoemission spectroscopy and inverse photoemission spectroscopy measurement^{53,54}. Comparing the difference among site-specific STS curves, it is evident that the density state of the center adatom contributes the most to the -0.8 V peak on the surface states. By employing statistical analysis considering data variance in big data, which is obtained by an AI-assisted measurement approach to expand the IV curve to a data set, our method has shown a demonstration of the room temperature measurement with reliability and validity.

DISCUSSION

Our research shows that the AI-SPM system has the potential to revolutionize room temperature SPM by addressing longstanding challenges and opening new avenues for materials characterization. This implementation can also be used in variable temperature STM as well as cryogenic SPM. The thermal correction will be less relevant for the cryogenic environment but the automation part can be used. In fact, even in cryogenic environments, in experiments over the span of days, thermal drift will occur and the tip of the probe will change. For site-specific experiments over a long period of time, AI-SPM will be definitely a powerful tool in cryogenic environments.

In room temperature environments, various heat-assisted processes, including diffusion, crystal growth, dislocation movements, and chemical reactions, occur at the nano- to atomic-scale. Precise site-specific measurements could potentially unveil new scientific domains if local conditions are accurately assessed. However, thermal drift and tip apex change in SPM often obstruct the precise measurements. Furthermore, the necessity to collect and statistically analyze extensive datasets in heat-present environments presents a substantial challenge, often beyond the capacity of manual operations. Previously, researchers have achieved excellent results with diffusion and atomic manipulation using room temperature SPM, but it required patience, time, and luck.

To overcome these difficulties at room temperature, we have shown that a deep learning-assisted AI-SPM system capable of autonomously collecting real-time data at room temperature can be deployed. A neural network is trained for comprehensive Si(111)-(7 × 7) topography assessment, achieving an impressive accuracy at around 90%, thanks to our two-phase data acquisition scheme which can automatically collect training datasets in real-time measurement. Moreover, this dataset acquisition routine can be applied to various surfaces, allowing users to train their own AI models for experimentation.

This transformation is evident not only in its capacity to assist humans in accomplishing complex and time-consuming operations, thus enabling automated experiments but also in the extension of SPM to the big data aiming to unlock deeper physical discoveries, as we delve into the analysis of a vast amount of data. In the future, it could be used effectively for SPM analysis at the atomic to nanometer level in materials with high temperatures and temperature variations, such as vanadium dioxide and thermoelectric materials. To achieve this vision, we need higher-bandwidth, automated SPM systems coupled with AI and real-time data analysis algorithms to provide intelligent data that is directly relevant to experimental conditions and scan results. However, there is a limitation of our current system, which demands preliminary experiments for executing the automatic data acquisition routine and gathering the training dataset necessary for the automation AI. In future developments,

training large models such as Vision Transformer⁵⁵, capable of encompassing a wide range of material structures would enable the utilization of neural networks for the analysis of general SPM data.

This manuscript has introduced the concept of a self-driving lab, setting the stage for the development of innovative approaches, such as the automation and optimization of manufacturing processes, and the operation, self-replication, and self-repair of molecular machines in the realm of atomic technology. These advancements align closely with Drexler’s anticipated integration of nanotechnology and AI¹.

CONCLUSIONS

We demonstrated the effectiveness of the AI-SPM system through two automated scanning experiments conducted at room temperature. We implemented an AI-equipped SPM capable of compensating for thermal drift and repairing tip damage, which have been challenges specific to the room-temperature operation. The system automatically addressed the challenge of room temperature by locating adsorbate-free regions and acquiring a large number of $I - V$ curves at four different adatom sites of the Si(111)-(7×7) surface. Our results indicated a 6.3% probability of changes occurring in the probe tip during the $I - V$ measurement process. This suggests that for reliable measurements at room temperature, it is necessary to acquire a significant amount of big data and subject these data to statistical processing. Such a process was not feasible with conventional SPM controllers for data acquisition, highlighting the need for AI support. STS measurements on the Si(111)-(7×7) demonstrated experimentally that the electronic state varies across four different adatom sites, confirming that STS characterization can be effectively performed at room temperature.

ACKNOWLEDGEMENTS

This work was supported in-part by a Grant-in-Aid for Scientific Research (19H05789, 21H01812, 22K18945) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT).

METHODS

Sample preparation and experiment environment

All experiments were carried out using a home-built STM operated at room temperature under ultrahigh vacuum conditions ($< \times 10^{-8}$ Pa). The experiments were repeated across multiple sessions with different Pt/Ir STM tips to ensure reproducibility. A n-type low-doped ($\rho \leq 0.02 \Omega\text{cm}$) Si substrate is used in this research. Atomically flat and clean Si(111)-(7×7) surfaces were

prepared with the standard cleaning procedure, and used for dataset acquisition and experiment demonstration.

For data acquisition, we have established SPM system augmented with a deep learning model explicitly designed for facilitating autonomous measurements, as depicted in Fig. 1. Our STM implementation combines an SPM instrument with a server that includes an “AI Inference” subsystem for deep learning prediction, and a real-time operating subsystem with “Scan Module” blocks that remotely control the SPM hardware [Fig. 1]. The control unit of the SPM instrument is built on an FPGA with remote access from a PC. The system was built with LabVIEW, LabVIEW FPGA, and Python. The scanning and data acquisition methods were performed in Python. NI PXIe-7857r was used as the measurement board. The Scan Module block in the server contains SPM automation functions for optimizing the measurement environment. It encompassed the scan operation with custom external scripts in Python that automated SPM measurement routines and contained scan functions that optimized the experimental environment. OpenCV⁵⁶ and SPMUtil⁵⁷ python packages were employed for data processing and image processing. Communication between the AI Inference subsystem and the SPM instrument exploited a TCP protocol connection. AI Inference subsystem runs on Python and PyTorch⁵⁸ is adapted as the machine learning framework to perform training and inference, which is accelerated by RTX 4090 GPU for tensor computation.

Thermal drift compensation

The real-time thermal drift compensation is based on an algorithm that extracts and matches feature points across consecutive scanned images⁴⁵. By computing pixel shifts between consecutive scan images, the scan area is offset to track the original region corresponding to the first image. Furthermore, the drift velocity along the x , y , and z is calculated from inter-image shifts and the data acquisition time and is utilized as the real-time drift compensation speed using a feedforward technique^{15,16}. The compensation process proceeded iteratively, acquiring images and adjusting for drift until the measured drift fell below a 0.2 nm threshold. Drift compensation is enabled to update the drift velocity at an interval of 10 min after the previous compensation is completed.

Tip apex optimization

A protocol of tip apex optimization is implemented to modify and evaluate tip apex conditions through controlled mechanical impacts³³. The bad tip is intentionally brought to poke toward the surface, inducing apex changes. As a model case, the probe is indented 0.9 nm towards the surface relative to the 1.5 V sample bias and 200 pA tunneling current setpoint. After poking,

the subsequent scanned image is input to Net1 for tip quality determination. Unsatisfactory tip states based on the network output triggered additional pokes. If a poke does not induce a change in the probe, the next poke will move the probe an additional 0.15 nm closer to the surface. In this work, the K_i ($i = 0, 1, 2, 3$) label in Fig. 2 (b) is regarded as a desired tip state for further experiments. The efficacy and responsibility of this automated optimization routine at room temperature have been demonstrated previously.

Training the deep learning model

Neural network types and hyperparameters for training Net1, Net2, and Net3 used in this work are listed in Tab. ???. Net1 uses a custom convolutional network structure which is presented in our previous study³³. Net2 and Net3 use the YOLOv8-small and YOLOv8-large structure⁵⁹ and the initial trainable parameters in the network are loaded from the pre-trained model. The YOLOv8-large in Net3 has more trainable parameters than YOLOv8-small and the larger model is chosen to improve the positioning accuracy on locating the atom key point.

At the first stage for training Net1, Net2, and Net3, input images were pre-processed with the plane fit subtraction and Gaussian-Hann filters, and the size of the filtered images were adjusted to 256 x 256 pixels. The training dataset can be automatically labeled by the trained model. Subsequently, the labeled dataset is exported to CVAT for manual validation. A balanced dataset was created, comprising more than 10,000 samples collected from experiments. Within this dataset, 2082 samples were designated for training Net1, 255 samples for training Net2, and 545 samples for training Net3. These datasets were divided into 80% for training and 20% for validation. Then, data augmentation⁶⁰, which contained random affine transform, image coping, and contrast change, was separately applied to the training dataset and the validation dataset and expands the dataset by 5 times. The three networks (Net1, Net2, and Net3) were trained by AdamW⁶¹ optimizer.

$I - V$ curves similarity metric and selection

To extract the high tendency from the part of the $I - V$ curves data group, the cosine similarity metric ϕ is used to process statics analysis. The similarity ϕ of two $I - V$ curves I_i, I_j ($i, j = 1, 2, \dots, m$) each containing n count data points in one curve is calculated by the following equation.

$$\phi(I_i, I_j) = \frac{\sum_{k=1}^n I_i[k] \cdot I_j[k]}{\sqrt{\sum_{k=1}^n I_i[k]^2 \cdot \sum_{k=1}^n I_j[k]^2}} \quad (1)$$

Especially, $\phi(I_1, I_2)$ is a range from -1 to 1, where $\phi(I_1, I_2) = 1$ means the two curve is fully equal. By calculating the ϕ among all $I - V$ curves, we can obtain a group of curves that can represent the most common data of the whole data. The major group selection method can be applied in two ways, which are based on the overall data (overall selection) or based on selected data (reference data-based selection), respectively. As for the overall selection method, given the whole $I - V$ curve data as $I_1, I_2, \dots, I_l, \dots, I_m$, we first calculate the mean value of cosine similarity ($\phi_{\text{mean}}[I_l]$) for each $I - V$ curve responding to the other all curves.

$$\phi_{\text{mean}}([I_l]) = \sum_m^i \phi([I_l, I_i]) / m \quad (2)$$

The $\phi_{\text{mean}}([I_l])$ for I_l means the common property to the whole $I - V$ curves, so we then apply a threshold T_1 with the condition $\phi_{\text{mean}}([I_l]) > T_1$ to choose all I_l curve that meets this condition as a major group. The reference data-based selection method is to select a group of data similar to the reference data, which is one of an $I - V$ curve from the $I - V$ curves group as the reference data I_{ref} . For an $I - V$ curve I_l , we give a threshold T_2 and use the $\phi(I_{\text{ref}}, I_l) > T_2$ condition to judge if I_l can be added into the major group which is based on I_{ref} . Before comparing the relationship between the $I - V$ curves, the measurement noise is reduced by the Savitzky-Golay⁵¹ and mean filters. After the major group is extracted, The most proper tendency in the group can be extracted by taking the average value in the $I - V$ curves and smoothing by a 1D spline function.

I. MODEL EVALUATION

Several metrics are introduced to assess model performance, specifically to validate the accuracy of classification and detection. For Net1, a confusion matrix is employed to verify classification accuracy as in Fig. 6(a). In the matrix, each element represents the number of instances by authors (horizontal axis) and the Net1 predictions (vertical axis), and it is normalized by the number of ground truth instances for each class. The diagonal elements represent the *Recall* metric, defined as:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

where true positive (TP) denotes the number of predictions identified correctly as positive among the actual positives, and false negative (FN) refers to the number of actual positives incorrectly predicted as negative. The Recall metric serves as an indicator of reproducibility when the model is applied within a measurement system. For instance, as illustrated by the first row and column in Fig. 6(a), scanning with a tip in the "Good 1" state provided the system with 0.98 likelihood of successful identification. The average *Recall* across all classes

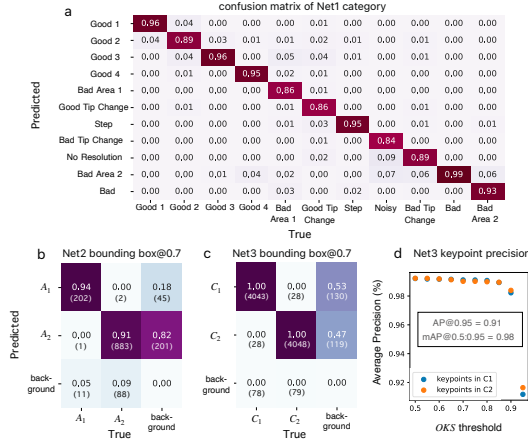


FIG. 6. Evaluation of the CNN performance of the room temperature STM on the Si(111) – (7 × 7). (a) to (c) Confusion matrix demonstrating classification accuracy of Net1 to Net3, respectively. In (b) and (c), background means AI does not treat the detection object as an interest to accurately identify, and value of $IoU \geq 0.7$ is defined as the true positive case. Numbers enclosed in brackets within the matrix elements, representing the number of test samples. (d) OKS threshold dependence of AP(Average Precision) analysis of Net3 for atom point detection. mAP calculates the mean value of the values of AP corresponding to the OKS from 0.5 to 0.95.

was 0.93, and the *Recall* for binary judgment, determining whether a tip was good or bad, was 0.98. Compared to our previous study³³ in which the same algorithm of Net1 has been used, we have confirmed the improvement of the *Recall* value of Net1 from 87% to 93%. This enhancement is attributed to the expansion of the dataset to 2000.

For Net2, the performance of the moving adsorption (A_1), and non-moving adsorption (A_2) detection is validated. To define the success or failure of the detection, the Intersection Over Union(IoU) metric which quantifies the overlap between the predicted bounding box and the ground truth bounding box for a given object instance, was introduced. It is defined as the intersection area between the predicted box and ground truth box divided by the union of the two boxes,

$$IoU = \frac{S_{\text{overlap}}}{S_{\text{union}}}, \quad (4)$$

where S_{overlap} is the area of overlap of the predicted and ground truth boxes, and S_{union} is their union area, which covered by both boxes. We defined the threshold value of $IoU = 0.7$ to determine if a prediction should have been considered a true or false positive. Specifically, a predicted box with $IoU \geq 0.7$ concerning the ground truth is counted as a true positive (TP), while a false negative (FN) indicates a failure to detect a ground truth object. Using this categorization, a confusion matrix can then be constructed to summarize the performance of each object class on the validation datasets. Fig. 6 (b) shows the

normalized confusion matrix when $IoU \geq 0.7$ was used as the threshold to divide the predictions in validation datasets. The category label also contains “background”, which represents the object that does not contain any target. The reason that a certain amount of background is misidentified as a target object may be because our detection model will repeatedly detect specific targets. The average Recall of detecting A_1 and A_2 is 0.92.

For Net3, Object Keypoint Similarity(OKS) was introduced to verify the detection accuracy of the bounding box (C_1 , C_2) and key points (P_i) as,

$$OKS = \frac{\sum_i^n \exp(-d_i^2/2s^2)\delta(v_i > 0)}{\sum_i^n \delta(v_i > 0)}, \quad (5)$$

where n is the number of key points, d_i is the Euclidean distance between the predicted keypoint and its corresponding ground truth, s is the region scale of the detected object, and $\delta(v_i > 0)$ is an indicator function that resolves to 1 if keypoint is visible or 0 if occluded. OKS ranges from 0 to 1, with 1 indicating perfect alignment of predicted key points to the ground truth. A threshold value of $OKS = 0.5$ was used to determine true versus false predicted key points and $OKS \geq 0.95$ can give a rigorous evaluation precision of the atom localization quality. The overall keypoint detection performance can then be quantified by the proportion of key points with OKS above a threshold. This is referred to as the mean Average Precision (mAP) for one of the keypoint classes.

In Fig. 6 (c), the normalized confusion matrix of two unit cells (C_1 , C_2) and background is shown. Here, similarly to Net2, $IoU = 0.7$ was chosen as the threshold to plot the normalized confusion matrix of two unit cells (C_1 , C_2) and background. A substantial quantity of unit cell training samples allowed the *Recall* value to approach its maximum performance level of 1.0.

Besides, the capability of accurately localizing atom keypoints can be evaluated by OKS value. When choosing OKS thresholds varying by 0.05 from 0.5 to 0.95, the respect of Average Precision detected in C_1 and C_2 are plotted in Fig. 6 (d). mAP was above 0.98, and as the most rigorous metric, mAP value at $OKS \geq 0.95$ was 0.91. Overall, these models are trained on the variety of the dataset which are acquired in real-time by a step-up data acquisition scheme. All of the models achieved high accuracy performance even for real-time data at room temperature, thus ensuring the robustness of the measurements and localization performance for SPM automation.

II. STATISTICAL METHODS IN ROOM TEMPERATURE $I - V$ MEASUREMENTS

Figure 7 shows the results of a long-time site-specific $I - V$ measurement, performed to verify the rate of change of the probe. The measurement is the same as in Fig. 5, but a different STM tip was used. The measurement took 58 hours and totally 695 topographic STM

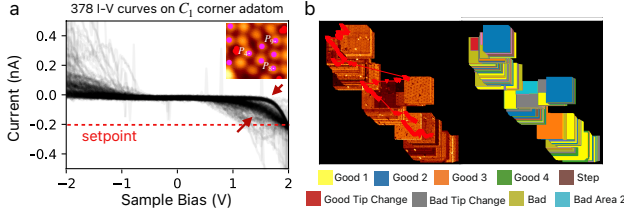


FIG. 7. Site-specific $I - V$ measurement by AI-SPM on the Si(111) - (7×7) surface at room temperature. The STM tip was repaired while searching the region, and data were automatically acquired in the region deemed appropriate for the $I - V$ measurement. (a) $I - V$ curves measured at the corner adatoms of faulted half-unit cells (C_1) on the Si(111) - (7×7) surface. A total of 378 curves are superimposed. The horizontal red dashed line indicates the set point of the tunneling current before the start of the IV scan. "P₄", "P₈", and "P₉" in the inset are corner adatoms within C_1 identified by Net3, where the $I - V$ curves were automatically measured. In reality, measurements were acquired at four different atomic sites on Si(111). (b) Top-most inferred state of the STM images during the self-driving $I - V$ measurements by our AI-SPM. The notation is the same as in Fig. 4(b).

images and 2832 $I - V$ curves were obtained. The selected 378 $I - V$ curves acquired on corner atoms of the half unit cell (the 4th, 8th, and 9th points detected by Net3) are shown in 7(a). Trajectory of the imaging area and the top-most inferred state of the STM images during the self-driving $I - V$ measurements by our AI-SPM is shown Fig. 7(b). The $I - V$ curve measurement contains both a forward sweep (I_{fw}), ranging from 2 V to -2 V of sample bias, and a backward sweep (I_{bw}) from -2 V to 2 V of sample bias. The scan is initiated from the target atom where the z position is specifically at a -200 pA setpoint on a 2 V sample bias. IV curves in Fig.8(a) encompass both I_{fw} and I_{bw} and almost all curves precisely align at the -200 pA setpoint under a 2 V sample bias position. This observation indicates that drift along the x , y , and z axes has been effectively corrected, offering compelling evidence that the system can mitigate the impact of thermal drift even during room temperature measurements.

However, due to various environment changes during the measurement process, the numerical values of I_{fw} and I_{bw} may not always align. In cases where the tip stabilizes, and the surface remains undisturbed, I_{fw} and I_{bw} tend to be consistent. To assess the level of agreement between I_{fw} and I_{bw} curves, we used the cosine similarity metric (see Method). In Fig. 8(a), as examples, three curves (iv-0, iv-1, and iv-2) are shown. As in the case of iv-0 and iv-1, respective ϕ values of 0.72 and 0.83 are evident that tip apex change happened. In contrast, when there is a high degree of consistency, as observed in the curve of "iv-2," the curve achieves a ϕ value of 1.00.

All $\phi(I_{fw}, I_{bw})$ values across a set of $I - V$ curves are computed and depicted in the histogram in Fig. 8(b). We defined $\phi(I_{fw}, I_{bw}) < 0.86$ as indicating a flaw in

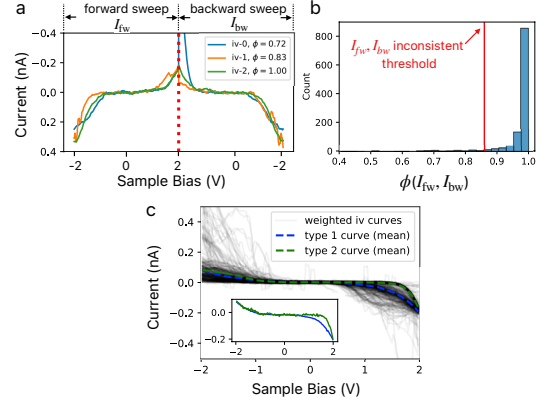


FIG. 8. Statistical analysis on 378 curves on C_1 corner adatom. (a) Three raw data("iv-0", "iv-1", "iv-2") of the IV measurement, which is taken from setting sample bias from 2 V to -2 V as the forward sweep and then -2 V to 2 V as the backward sweep, starting from -200 pA setpoint at 2 V sample bias. (b) Histogram of the similarity metric $\phi(I_{fw}, I_{bw})$ between forward $I - V$ curve I_{fw} and backward $I - V$ curve I_{bw} . (c) The result of two $I - V$ curves which represent the two tend group's mean value. Type 1 curve is selected by the overall data-based selection method and type 2 curve is selected by the reference curve-based selection method (see Methods).

$I - V$ measurements caused by changes in measurement conditions, the probability of encountering such measurement discrepancies using our STS parameter was 6.3%. The reason behind these discrepancies is often associated with the impact on the tip apex resulting from the applied sample bias, which can be likened to a process of conditioning the probe²⁹. These cases are frequently observed during measurements, manifesting as tip changes or impurities dropping onto the surface, and are largely unavoidable when operating at room temperature.

Despite the potential influence of an unstable measurement environment as described earlier, trends in surface properties can still be observed, especially when supported by adequate data. Considering that all the curves have been weighted in the plot, the areas with denser line colors indicate higher-weight regions, reflecting a more pronounced tendency in the $I - V$ curve within those specific regions. Within the range of 1 to 2 V Sample Bias, two $I - V$ curves exhibit the most prominent directions. To distinguish and represent these two distinct trends, an algorithm based on cosine similarity (see Method) is employed to statistically differentiate and select the sets of $I - V$ curves representing these two tendencies. The average values of all curves within each set, corresponding to the two tendencies, are plotted as blue and green dotted lines in Fig. 8(c). The reason behind the differentiation of these two trends on corner adatoms of the C_1 type, despite their identical nature, might be attributed to variations in the density of the state of the STM tip apex during $I - V$ curve measurements. Though our automated SPM system guarantees the ability to ac-

quire atomic-resolution topography, it still needs to be improved to maintain consistent local density of states of the tip during measurements.

- ¹K. E. Drexler, *Engines of Creation: The Coming Era of Nanotechnology* (Anchor Press/Doubleday, Garden City, NY, 1986).
- ²G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, *Phys. Rev. Lett.* **49**, 57–61 (1982).
- ³G. Binnig, C. F. Quate, and C. Gerber, *Phys. Rev. Lett.* **56**, 930–933 (1986).
- ⁴D. M. Eigler and E. K. Schweizer, “Positioning single atoms with a scanning tunnelling microscope,” *Nature* **344**, 524–526 (1990).
- ⁵P. Jelínek, *Journal of Physics: Condensed Matter* **29**, 343002 (2017).
- ⁶C. Barth and M. Reichling, *Nature* **414**, 54–57 (2001).
- ⁷N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando, *Nature* **468**, 72–76 (2010).
- ⁸V. Jelic, K. Iwaszczuk, P. H. Nguyen, C. Rathje, G. J. Hornig, H. M. Sharum, J. R. Hoffman, M. R. Freeman, and F. A. Hegmann, “Ultrafast terahertz control of extreme tunnel currents through single atoms on a silicon surface,” *Nature Physics* **13**, 591–598 (2017).
- ⁹S. Yoshida, Y. Terada, R. Oshima, O. Takeuchi, and H. Shigekawa, “Nanoscale probing of transient carrier dynamics modulated in a gaas–pin junction by laser-combined scanning tunneling microscopy,” *Nanoscale* **4**, 757–761 (2012).
- ¹⁰I. Brihuega, O. Custance, and J. M. Gómez-Rodríguez, *Phys. Rev. B* **70**, 165410 (2004).
- ¹¹O. Custance, I. Brihuega, J. Gómez-Rodríguez, and A. Baró, *Surface Science* **482–485**, 1406–1412 (2001).
- ¹²M. Meusel, A. Gezmis, S. Jaekel, M. Lexow, A. Bayer, F. Maier, and H.-P. Steinrück, *The Journal of Physical Chemistry C* **125**, 20439–20449 (2021).
- ¹³Y. Sugimoto, M. Abe, S. Hirayama, N. Oyabu, Ó. Custance, and S. Morita, *Nature Materials* **4**, 156–159 (2005).
- ¹⁴Y. Sugimoto, P. Pou, O. Custance, P. Jelínek, M. Abe, R. Pérez, and S. Morita, *Science* **322**, 413–417 (2008), <https://www.science.org/doi/pdf/10.1126/science.1160601>.
- ¹⁵M. Abe, Y. Sugimoto, O. Custance, and S. Morita, “Room-temperature reproducible spatial force spectroscopy using atom-tracking technique,” *Applied Physics Letters* **87**, 173503 (2005).
- ¹⁶P. Rahe, J. Schütte, W. Schniederberend, M. Reichling, M. Abe, Y. Sugimoto, and A. Kühnle, “Flexible drift-compensation system for precise 3D force mapping in severe drift environments,” *Review of Scientific Instruments* **82**, 063704 (2011).
- ¹⁷O. Albrektsen, H. W. M. Salemink, K. A. Mo/rch, and A. R. Thölen, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* **12**, 3187–3190 (1994).
- ¹⁸P. Hapala, G. Kichin, C. Wagner, F. S. Tautz, R. Temirov, and P. Jelínek, *Phys. Rev. B* **90**, 085421 (2014).
- ¹⁹B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette, “Self-driving laboratory for accelerated discovery of thin-film materials,” *Science Advances* **6**, eaaz8867 (2020).
- ²⁰M. Abolhasani and E. Kumacheva, “The rise of self-driving labs in chemical and materials sciences,” *Nature Synthesis* **2**, 483–492 (2023).
- ²¹K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” (2016).
- ²²S. Liu and W. Deng, (2015).
- ²³O. M. Gordon and P. J. Moriarty, *Machine Learning: Science and Technology* **1**, 023001 (2020).
- ²⁴S. Farley, J. E. A. Hodgkinson, O. M. Gordon, J. Turner, A. Soltoggio, P. J. Moriarty, and E. Hunsicker, “Improving the segmentation of scanning probe microscope images using convolutional neural networks,” *Machine Learning: Science and Technology* **2**, 015015 (2020).
- ²⁵O. M. Gordon, J. E. A. Hodgkinson, S. M. Farley, E. L. Hunsicker, and P. J. Moriarty, *Nano Letters* **20**, 7688–7693 (2020).
- ²⁶L. R. Hofer, M. Krstajić, P. Juhász, A. L. Marchant, and R. P. Smith, “Atom cloud detection and segmentation using a deep neural network,” *Machine Learning: Science and Technology* **2**, 045008 (2021).
- ²⁷R. Lin, R. Zhang, C. Wang, X.-Q. Yang, and H. L. Xin, *Scientific Reports* **11**, 5386 (2021).
- ²⁸S.-H. Yang, W. Choi, B. W. Cho, F. O.-T. Agyapong-Fordjour, S. Park, S. J. Yun, H.-J. Kim, Y.-K. Han, Y. H. Lee, K. K. Kim, and Y.-M. Kim, *Advanced Science* **8**, 2101099 (2021).
- ²⁹S. Wang, J. Zhu, R. Blackwell, and F. R. Fischer, “Automated tip conditioning for scanning tunneling spectroscopy,” *The Journal of Physical Chemistry A* **125**, 1384–1390 (2021), pMID: 33560124, <https://doi.org/10.1021/acs.jpca.0c10731>.
- ³⁰A. Krull, P. Hirsch, C. Rother, A. Schiffrin, and C. Krull, *Communications Physics* **3**, 54 (2020).
- ³¹J. C. Thomas, A. Rossi, D. Smalley, L. Francaviglia, Z. Yu, T. Zhang, S. Kumari, J. A. Robinson, M. Terrones, M. Ishigami, E. Rotenberg, E. S. Barnard, A. Raja, E. Wong, D. F. Ogle-tree, M. M. Noack, and A. Weber-Bargioni, *npj Computational Materials* **8**, 99 (2022).
- ³²I.-J. Chen, M. Aapro, A. Kipnis, A. Ilin, P. Liljeroth, and A. S. Foster, “Precise atom manipulation through deep reinforcement learning,” *Nature Communications* **13**, 7499 (2022).
- ³³Z. Diao, L. Hou, and M. Abe, “Probe conditioning via convolution neural network for scanning probe microscopy automation,” *Applied Physics Express* **16**, 085002 (2023).
- ³⁴Y. Sugimoto, A. Yurtsever, M. Abe, S. Morita, M. Ondráček, P. Pou, R. Pérez, and P. Jelínek, “Role of tip chemical reactivity on atom manipulation process in dynamic force microscopy,” *ACS Nano* **7**, 7370–7376 (2013), pMID: 23906095, <https://doi.org/10.1021/nn403097p>.
- ³⁵Y. Sugimoto, K. Miki, M. Abe, and S. Morita, “Statistics of lateral atom manipulation by atomic force microscopy at room temperature,” *Phys. Rev. B* **78**, 205305 (2008).
- ³⁶F. Ming, K. Wang, S. Pan, J. Liu, X. Zhang, J. Yang, and X. Xiao, “Assembling and disassembling ag clusters on si(111)-(7×7) by vertical atomic manipulation,” *ACS Nano* **5**, 7608–7616 (2011), pMID: 21819127.
- ³⁷Y. Sugimoto, A. Yurtsever, N. Hirayama, M. Abe, and S. Morita, “Mechanical gate control for atom-by-atom cluster assembly with scanning probe microscopy,” *Nature Communications* **5**, 4360 (2014).
- ³⁸E. Inami, I. Hamada, K. Ueda, M. Abe, S. Morita, and Y. Sugimoto, “Room-temperature-concerted switch made of a binary atom cluster,” *Nature Communications* **6**, 6231 (2015).
- ³⁹I.-S. Hwang, M.-S. Ho, and T. T. Tsong, “Dynamic behavior of si magic clusters on si(111) surfaces,” *Physical Review Letters* **83**, 120–123 (1999).
- ⁴⁰C. Zhang, G. Chen, K. Wang, H. Yang, T. Su, C. T. Chan, M. M. T. Loy, and X. Xiao, “Experimental and theoretical investigation of single cu, ag, and au atoms adsorbed on Si(111)-(7×7),” *Phys. Rev. Lett.* **94**, 176104 (2005).
- ⁴¹I. Öst’adal, P. Kocán, P. Sobotík, and J. Pudl, “Direct observation of long-range assisted formation of ag clusters on Si(111)7×7,” *Phys. Rev. Lett.* **95**, 146101 (2005).
- ⁴²K. Wang, G. Chen, C. Zhang, M. M. T. Loy, and X. Xiao, “Intermixing of intrabasin and interbasin diffusion of a single ag atom on Si(111)-(7×7),” *Phys. Rev. Lett.* **101**, 266107 (2008).
- ⁴³J. R. Osiecki, S. Suto, and A. Chutia, “Periodic corner holes on the si(111)-7×7 surface can trap silver atoms,” *Nature Communications* **13**, 2973 (2022).
- ⁴⁴O. Custance, I. Brihuega, J. Gómez-Rodríguez, and A. Baró, “Initial stages of sn adsorption on si(111)-(7×7),” *Surface Science* **482–485**, 1406–1412 (2001).
- ⁴⁵Z. Diao, K. Ueda, L. Hou, H. Yamashita, O. Custance, and M. Abe, “Automatic drift compensation for nanoscale imag-

- ing using feature point matching,” *Applied Physics Letters* **122** (2023), 10.1063/5.0139330, 121601.
- ⁴⁶R. J. Hamers, R. M. Tromp, and J. E. Demuth, “Surface electronic structure of si (111)-(7×7) resolved in real space,” *Phys. Rev. Lett.* **56**, 1972–1975 (1986).
- ⁴⁷H. Neddermeyer and S. Tosch, “Scanning tunneling spectroscopy on si,” *Ultramicroscopy* **25**, 135–147 (1988).
- ⁴⁸P. Avouris and R. Wolkow, “Atom-resolved surface chemistry studied by scanning tunneling microscopy and spectroscopy,” *Phys. Rev. B* **39**, 5091–5100 (1989).
- ⁴⁹J. Mysliveček, A. Stróżecka, J. Steffl, P. Sobotík, I. Ošťádal, and B. Voigtländer, *Phys. Rev. B* **73**, 161302 (2006).
- ⁵⁰A. B. Odobescu and S. V. Zaitsev-Zotov, “Energy gap revealed by low-temperature scanning–tunnelling spectroscopy of the si(111)-7×7 surface in illuminated slightly doped crystals,” *Journal of Physics: Condensed Matter* **24**, 395003 (2012).
- ⁵¹J. Steinier, Y. Termonia, and J. Deltour, “Smoothing and differentiation of data by simplified least square procedure,” *Analytical Chemistry* **44**, 1906–1909 (1972).
- ⁵²R. M. Tromp, R. J. Hamers, and J. E. Demuth, *Phys. Rev. B* **34**, 1388–1391 (1986).
- ⁵³F. J. Himpsel and T. Fauster, “Probing valence states with photoemission and inverse photoemission,” *Journal of Vacuum Science & Technology A* **2**, 815–821 (1984).
- ⁵⁴F. J. Himpsel, T. Fauster, and G. Hollinger, “Electronic structure of si(111) surfaces,” *Surface Science* **132**, 22–30 (1983).
- ⁵⁵A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” (2021), arXiv:2010.11929 [cs.CV].
- ⁵⁶G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools* (2000).
- ⁵⁷“Spmutil,” (2023), <https://github.com/DIAOZHUO/SPMUtil>.
- ⁵⁸A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” (2019).
- ⁵⁹G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” (2023).
- ⁶⁰A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, “Albumentations: Fast and flexible image augmentations,” *Information* **11** (2020), 10.3390/info11020125.
- ⁶¹I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” (2019).