

# Automated Tip Conditioning for Scanning Tunneling Spectroscopy

*Shenkai Wang,<sup>†</sup> Junmian Zhu,<sup>†</sup> Raymond Blackwell,<sup>†,‡</sup> and Felix R. Fischer<sup>\*,†,‡,§</sup>*

<sup>†</sup>Department of Chemistry, University of California, Berkeley, California 94720, USA.

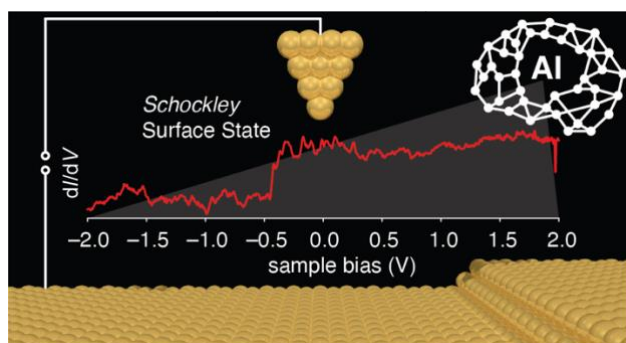
<sup>‡</sup>Kavli Energy NanoSciences Institute at the University of California Berkeley and the Lawrence  
Berkeley National Laboratory, Berkeley, California 94720, USA.

<sup>§</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California  
94720, USA.

**Corresponding Author**

\*Email: [ffischer@berkeley.edu](mailto:ffischer@berkeley.edu).

ABSTRACT. Scanning tunneling spectroscopy (STS), a technique that records the change in the tunneling current as function of the bias ( $dI/dV$ ) across the gap between a tip and the sample, is a powerful tool to characterize the electronic structure of single molecules and nanomaterials. While performing STS, the structure and condition of the scanning probe microscopy (SPM) tips are critical for reliably obtaining high quality point spectra. Here, we present an automated program based on machine learning models that can identify the Au(111) *Shockley* surface state in  $dI/dV$  point spectra and perform tip conditioning on clean or sparsely covered gold surfaces with minimal user intervention. We employ a straightforward height-based segmentation algorithm to analyze STM topographic images to identify tip conditioning positions and used 1789 archived  $dI/dV$  spectra to train machine learning models that can ascertain the condition of the tip by evaluating the quality of the spectroscopic data. Decision tree based ensemble and boosting models, and deep neural networks (DNNs) have been shown to reliably identify tips in suitable conditions for STS. We expect the automated program to reduce operational costs and time, increase reproducibility in surface science studies, and accelerate the discovery and characterization of novel nanomaterials by STM. The strategies presented in this paper can readily be adapted to STM tip conditioning on a wide variety of other common substrates.



TOC GRAPHICS:

**KEYWORDS.** scanning tunneling microscopy, scanning tunneling spectroscopy, automated tip conditioning, Python, machine learning

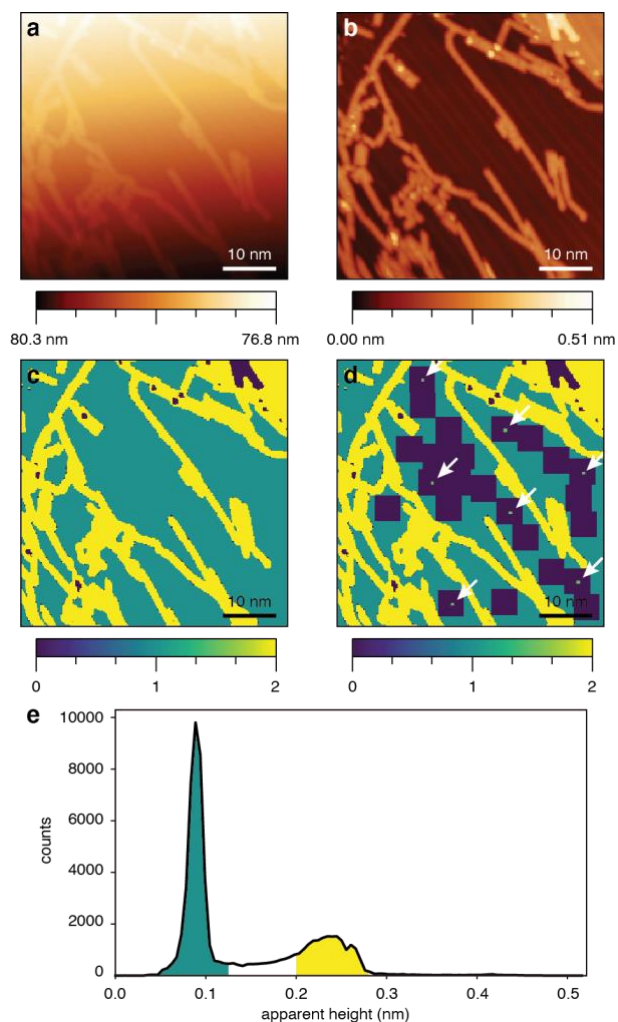
Scanning tunneling microscopy (STM) techniques<sup>1-3</sup> and associated spectroscopic (STS) methods,<sup>4-6</sup> such as  $dI/dV$  point spectroscopy, have been widely used to measure electronic structures and local density of states (LDOS) of single molecules and materials with unprecedented spatial and energy resolutions.<sup>7-9</sup> However, the quality of  $dI/dV$  spectra highly depends on the shape of the probe tips, and atomically sharp tips with stable, well-defined apex structures are required for obtaining reliable spectra.<sup>10-12</sup> In most cases, STS measurements are performed in ultra-high vacuum (UHV) and low temperature (4 K) to minimize external disturbances. Initial tip preparation<sup>13-15</sup> and continuous *in situ* tip conditioning<sup>15,16</sup> are often required throughout the characterization of target molecules and materials. A common way to prepare STM tip is to repetitively poke the tip into well-defined and bare substrates, *i.e.* coinage metals or silicon, to remove contaminations and in the process coat the tip with a thin layer of substrate atoms.<sup>17-19</sup> The  $dI/dV$  spectrum of the substrate will then be used as a reference to determine whether the tip is suitable for STS experiments. Since the tip geometry change during the poking process is unpredictable, the tip conditioning is typically slow and needs to be monitored continuously. It therefore represents a key factor that restricts the speed of acquisition of high-quality STM spectroscopic data. In order to make efficient use of the idle time of the instrument and minimize the operational time spend on tip conditioning, we developed a program based on Python and machine learning that can fully automate the time-consuming tip conditioning processes. The program is designed to perform tip conditioning on Au(111) surfaces that are clean or feature a sub-monolayer molecular coverage with minimal operator intervention. Once initiated, the program autonomously performs tip conditioning cycles until the tip can generate publication quality spectroscopic data on gold surfaces (see Supporting Information for the user interface). The automation package seamlessly interfaces with control software of a Scienta Omicron STM

and autonomously analyzes the collected topographic images to find flat bare regions of Au that are large enough for tip conditioning. The program proceeds to collect  $dI/dV$  spectra at automatically selected positions and uses machine learning models to compare their quality to standard  $dI/dV$  spectra for Au<sup>20</sup> to decide if the tip meets the requirements for further STS measurements. If the tip condition is inadequate, the program instructs the STM control software to continue the conditioning cycle until the machine learning model determines the quality of the tip to meet the set STS standard. We used 1789  $dI/dV$  spectra previously collected in our group to train machine learning models and to evaluate their performances. Decision tree based ensemble and boosting models (e.g. Random Forest, AdaBoost,<sup>21</sup> and CatBoost<sup>22</sup>) and deep neural networks (DNNs) show reasonable performance in identifying tips that are in usable conditions with clear Au(111) *Shockley* surface states (gold surface states) evident in the spectra. Experimental high-quality  $dI/dV$  spectra are rare and their numbers are inadequate to generate a sufficiently large training set to train a DNN. We demonstrate that data augmentation strategies for generating artificial  $dI/dV$  spectra based on experimental data can be used to solve the data availability issue. The data processing strategies presented here are generally applicable for automatic STM tip conditioning on other surfaces routinely used in STM experiments.

**Locating tip conditioning positions in topographic images.** In order for the program to perform autonomous tip conditioning, it needs to be able to identify suitable positions to poke the tip and record  $dI/dV$  point spectra. An algorithm that can process STM topographic images and identify unobstructed clean substrate areas is required. Programs aiming to automate topographic image analysis and STM operations have been reported before,<sup>19,23-30</sup> but most of these algorithms, some associated with machine learning methods, are designed for more complicated purposes such as atom manipulation. Our program employs a fast and direct method that segments topographic

images based on apparent height and scans the labeled images as a 5 nm by 5 nm window to locate possible substrate areas for tip conditioning. The image processing procedure is visualized in Figure 1. Topographic STM images were collected on a Au(111) substrate covered with a submonolayer of graphene nanoribbons (GNRs). Raw topographic images tend to be tilted, as shown in Figure 1a, since STM tips are generally not aligned perfectly perpendicular to the sample surface. Therefore, the first step of image processing is to calculate the best fit plane of the image using normal equation, a commonly used method in linear algebra, and subtract the plane from the raw image. The normal equation flattened image of Figure 1a is shown in Figure 1b. The program will analyze the apparent height distribution of the flattened image, visualized as a histogram in Figure 1c, and identify peaks that represent the most prevalent z-heights. The flattened image can thus be segmented based on apparent height and labeled by the sequence of peaks in the histogram, as shown in Figure 1d and highlighted in Figure 1e. For our GNRs/Au(111) systems or for clean Au(111) surfaces, we group pixels within  $-0.05$  nm to  $+0.05$  nm of each prevalent height together and assign them with the same label (Figure 1e). For pixels that are not included in any group, we label them 0 (shown as dark blue areas in Figure 1d and 1e). This range can be adjusted based on the height variances of different surfaces used in STM. After labeling the topographic images, the program will use a 5 nm by 5 nm window to scan the image line by line and register areas that contain the same non-zero label (Figure 1f). The middle of each registered square is considered a good location to poke or record  $dI/dV$  point spectra. This selection ensures that both the tip preparation and STS point spectroscopy will not be performed over molecular adsorbates or step edges. In each tip conditioning cycle, the program will control the STM to take two consecutive  $dI/dV$  spectra at one of the selected locations and use a machine learning model (discussed below) to analyze the second spectra. If the machine learning model predicts that the tip is not in the

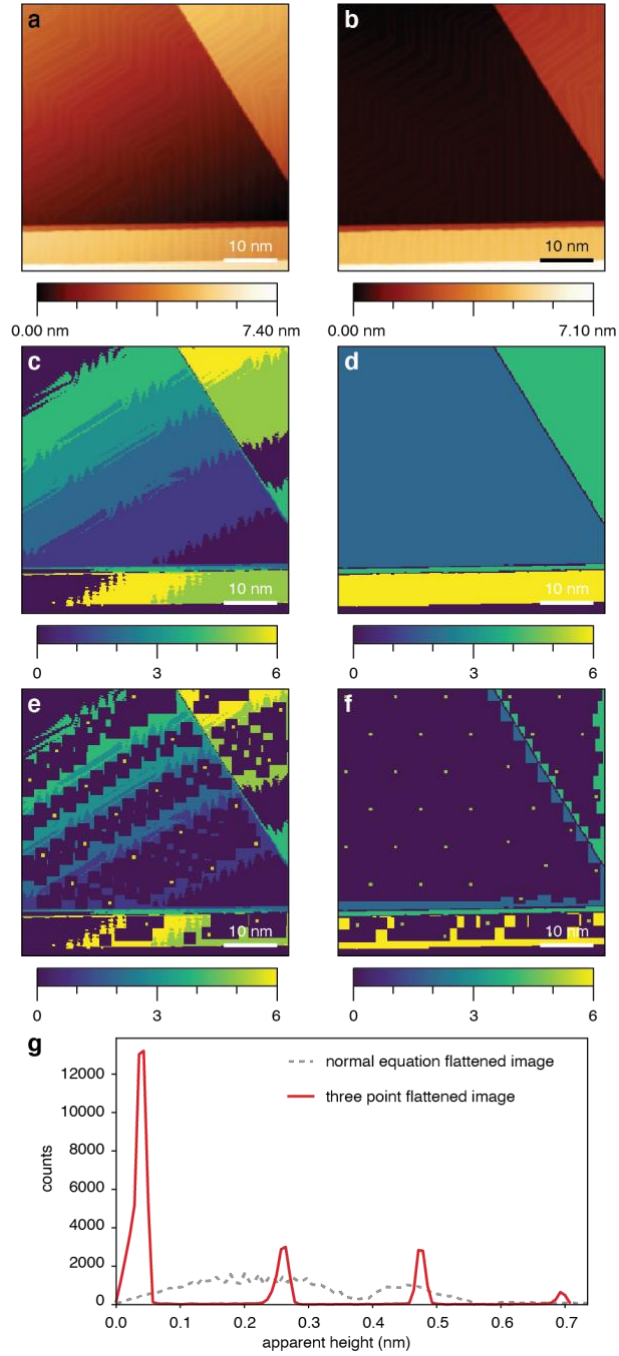
desired condition, the program will control the STM to perform a 2 nm poke at the same location and move to the next cycle. Tip conditioning cycles are performed at positions that are separated by at least 15 nm to avoid poking or performing STS over spurious particles dropped on the surface during previous conditioning cycles. In the example illustrated in Figure 1, the tip conditioning cycles are performed at locations marked by arrows in Figure 1d. The 5 nm window and 15 nm separations were determined based on the optimal requirements of our specific STM instrument but can be adjusted to other systems. We use machine learning models to analyze the second spectra in each tip conditioning cycle since the tip condition tends to change during the STS measurement immediately after a poke. This cycle is designed to be conservative to avoid potential damage to the tip and can be adjusted based on user preference.



**Figure 1.** Topographic STM image segmented and labeled by the automated program. (a) Raw topographic STM image of graphene nanoribbons on Au(111). (b) Figure (a) flattened using normal equation. (c) Figure (b) segmented and labeled topographic image based on apparent height (see text). (d) Image showing tip conditioning positions determined autonomously by the program. Arrows highlight positions where tip conditioning cycles (see text) will be performed (e) Histogram of pixel heights for Figure (b). Highlighted regions represent  $-0.05$  nm to  $+0.05$  nm range centered around each peak. Figure (c) is labeled based on the highlighted regions.

Conventional strategies used to flatten topographic images using normal equation will not completely level the image if molecules on the surface are not evenly distributed or if step edges

are present within the scanning window. To resolve these cases, STM users normally choose three points in the image that are expected to have the same height to define a new best fit plane for image flattening. In our program, we incorporate a function that can perform three point flattening automatically. After the program identifies all suitable positions for tip conditioning using the method mentioned above, it will pick the central coordinates of the three most separated squares in the label 1 region. The average height within each of the three squares in the topographic image is assigned to the corresponding central coordinate and used to calculate the new best fit plane. To demonstrate the performance of automatic three points flattening, we present an example in Figure 2. Figure 2a,c,e show the results of an STM image flattened by normal equation and labeled using the methods demonstrated in Figure 1. Figure 2b shows the three points flattened result of Figure 2a. The height distributions of Figure 2a and 2b are presented in Figure 2g, from which we can see that the additional three points flattening technique leads to a significant improvement in the leveling of the image when compared to the normal equation method. Furthermore, our segmentation and labeling algorithm performs much better on three point flattened images, as demonstrated in Figure 2d,f.

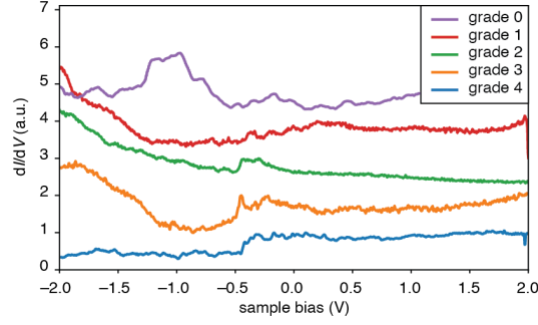


**Figure 2.** Topographic STM image processed with different methods by the automated program.

(a) STM image of Au(111) surface flattened using normal equations. (b) Figure (a) flattened using three points alignment. (c) Figure (a) segmented and labeled based on apparent height. (d) Figure (b) segmented and labeled based on apparent height. (e) Image showing tip conditioning positions

detected by the program in Figure (a). (f) Image showing tip conditioning positions detected by the program in Figure (b). (g) Histogram of pixel height for Figure (a) and (b).

**Evaluating tip condition from  $dI/dV$  spectra using machine learning models.** For a tip conditioning program to be practical, it needs to be able to exit the cycle when the tip is in a state suitable for STS measurement. The program is designed to work on clean or sparsely covered Au(111) surfaces, so the quality of  $dI/dV$  spectra taken on gold substrate itself can be used as an indicator for the quality of the tip. While an experienced researcher can readily identify a clean STS spectrum recorded e.g. on Au(111), there is no unique criteria that could be hard coded into an automated program. To resolve this challenge, we herein apply machine learning models trained on a library of archived gold  $dI/dV$  spectra to analyze and evaluate the traces collected as part of the tip conditioning process. We manually labeled archived spectra based on parameters like signal to noise ratio, position of the characteristic surface state, and appearance of unknown features with respect to an idealized  $dI/dV$  point spectrum of a Au(111) surface. Each  $dI/dV$  spectrum is graded on a scale between 0 to 4, with 4 corresponding to an ideal curve and 0 to curves entirely lacking a defined feature for the gold surface states or featuring significant undefined peaks from tip adsorbates (including spectra taken on molecules or on other substrates). A sample of  $dI/dV$  spectra with different grades are presented in Figure 3. Spectra with clear gold surface states are graded between 2 to 4 while spectra without gold surface states are graded as 0 or 1. The number of  $dI/dV$  spectra with different grades are listed in Table 1. As our archive only contains a limited number of spectra graded at 3 and 4 for training, we initially grouped  $dI/dV$  spectra with grade 2, 3, and 4 together and train them against spectra with grade 0 and 1 to obtain models that can reliably identify a *Schockley* gold surface state.



**Figure 3.** Example STS curves that have been assigned with different grades.

**Table 1.** Number of  $dI/dV$  spectra with different grade. Spectra with clear gold surface states are graded between 2 to 4.

| Grade                                         | 0(Bad) | 1   | 2(OK) | 3  | 4(Good) | Total |
|-----------------------------------------------|--------|-----|-------|----|---------|-------|
| <b>Training Label<br/>(Au Surface State?)</b> | False  |     | True  |    |         |       |
| <b>Number</b>                                 | 410    | 955 | 334   | 84 | 6       | 1789  |

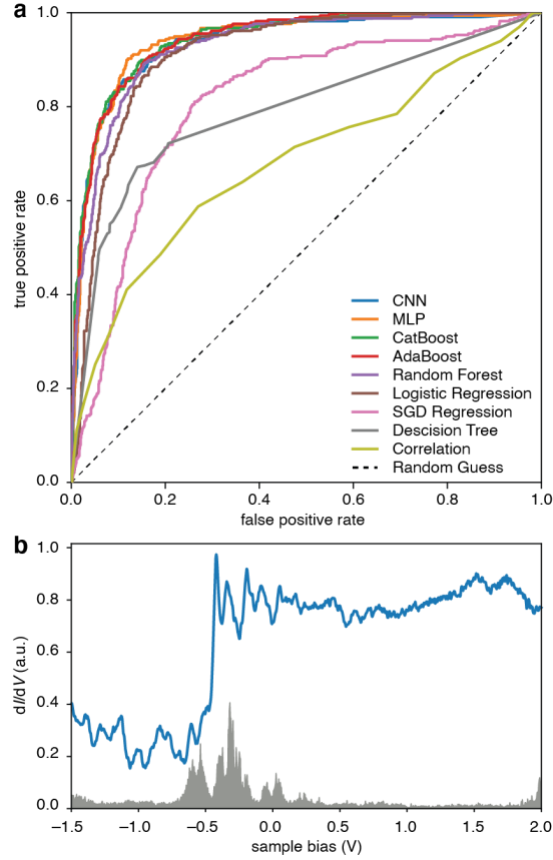
Before training machine learning models, we uniformly scale the STS data by normalizing the maximum  $dI/dV$  value between  $-0.75$  V and  $0.00$  V in each curve to 1, so that the relative intensity of the gold surface state can readily be determined during the normalization. We note that  $-1.5$  V to  $2.0$  V represents the most important bias window in STS measurements, so in each  $dI/dV$  spectrum we pick  $dI/dV$  values at 896 evenly spaced biases between  $-1.5$  V to  $2.0$  V and generate a new curve suitable for machine learning training and prediction. We used 80% of the archived data (1094 labeled *false* and 338 labeled *true*) for training and 20% of the data (271 labeled *false* and 86 labeled *true*) for testing. To evaluate the performance of different models, we compared their precision scores and recall scores acquired on the test set and their Receiver Operating Characteristic (ROC) curves generated from three-fold cross-validation data on a training set.<sup>31</sup> To benchmark the performance of our machine learning models we used a simple correlation model

that does not rely on machine learning as a reference. This correlation method classifies the spectra in our data set based on their mean squared error with respect to a highest quality reference  $dI/dV$  spectrum (see Supporting Information for details). The results of our benchmarking results are summarized in Table 2 along with the corresponding ROC curves in Figure 4a.

**Table 2.** Performance of different machine learning models\* on differentiating STS curves with gold surface states.

| <b>Model</b>                   | <b>Precision (On test set)</b> | <b>Recall (On test set)</b> | <b>ROC<br/>Area Under Curve</b> |
|--------------------------------|--------------------------------|-----------------------------|---------------------------------|
| <b>Correlation</b>             | 0.408                          | 0.529                       | 0.679                           |
| <b>SGD</b>                     | 0.595                          | 0.547                       | 0.807                           |
| <b>SVM</b>                     | 0.685                          | 0.733                       | N/A**                           |
| <b>Logistic<br/>Regression</b> | 0.587                          | 0.628                       | 0.913                           |
| <b>Decision Tree</b>           | 0.746                          | 0.616                       | 0.790                           |
| <b>Random Forest</b>           | 0.825                          | 0.605                       | 0.928                           |
| <b>AdaBoost</b>                | 0.829                          | 0.674                       | 0.942                           |
| <b>CatBoost</b>                | 0.842                          | 0.744                       | 0.943                           |
| <b>MLP</b>                     | 0.792                          | 0.663                       | 0.940                           |
| <b>CNN</b>                     | 0.806                          | 0.674                       | 0.935                           |

\*SGD, SVM, MLP, and CNN denote stochastic gradient descent, support vector machine, multilayer perceptron, and convolutional neural network, respectively. AdaBoost and CatBoost are boosting methods based on decision tree. Correlation represents a basic model that does not rely on machine learning (see text). See supporting information for model details. \*\*SVM is generally not used to predict the probability of classification so a ROC curve is not applicable for SVM.



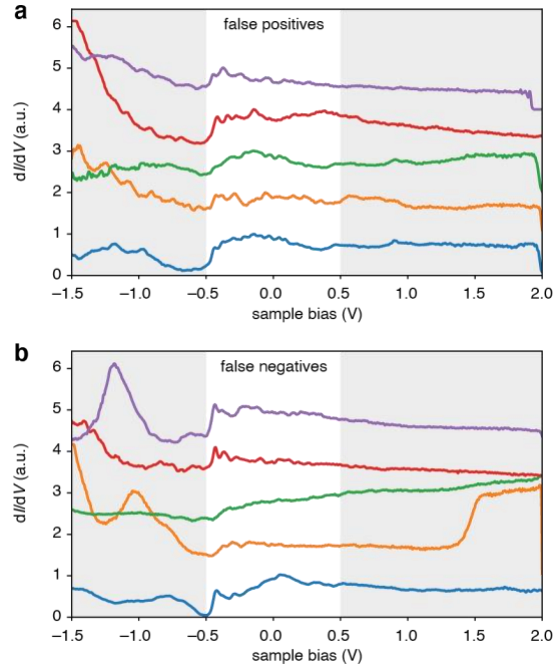
**Figure 4.** Performance of machine learning models on differentiating STS curves with gold surface states. (a) Receiver Operating Characteristic (ROC) curves for machine learning models. The curves are generated from classification probability results of three-fold cross-validation on the training set (except for the correlation curve, see Supporting Information). (b) Contribution of each data point on the classification of STS curves using a random forest model (feature importance). An STS curve with grade 4 is presented as reference in blue.

From Table 2 and Figure 4a we can see that the correlation method, regression methods, and several basic machine learning methods can in general differentiate high quality STS curves from poor ones but their classification results are inferior to decision tree based ensemble and boosting models or deep neural networks (MLP, CNN). Figure 4b shows a representation of the weight of each data point in the STS curve on making the classification decision if a random forest model is

applied. We notice that the model performs classification mainly based on data points between  $-0.75$  V and  $0.00$  V, emphasizing the characteristic region associated with the gold surface state. This suggests that machine learning models can indeed learn to differentiate a gold surface state based on our labeling.

While the performance of machine learning models judged just by the precision and recall scores alone (Table 2) is consistently superior to correlation and regression models the identification of high-quality STS spectra is not flawless. We largely attribute this to the inconsistency of manual labeling of the training set since there is no objective criteria for grading the  $dI/dV$  spectra. The differences between spectra with different grades are often minor, especially for spectra that are graded as 1 or 2. Therefore, ambiguity in manual classification introduces significant noises into the dataset. The limited number of grade 4 spectra for training is another reason for the performance of machine learning models. Deep neural networks tend to overfit our sample despite the simple architectures (see Supporting Information) we used in our program. Furthermore, for the machine learning models to work we have to group spectra with grade 2, 3, and 4 together and label them as acceptable for STS. In order to better understand the qualitative performance of machine learning models, we presented in Figure 5 some of the  $dI/dV$  spectra in the test set that were predicted as *false positives* and *false negatives* by an AdaBoost model. We can see from Figure 5a that the spectra that have been predicted as *false positives* are not entirely unsuitable for STS, and if the tip conditioning program were to stop when obtaining one of these spectra, the operator may only need to do a few more mild tip conditioning operations before applying the tip to experimental measurements. On the other hand, the *false negative* spectra presented in Figure 5b are not ideal gold spectra and represent tips that show undesirable artifacts in some sections of the spectrum. These false negative spectra may have been graded as 2 and labeled as *true* for

training. Therefore, we are satisfied with the performances of our AdaBoost, CatBoost models and deep neural networks on identifying STM tips that are in reasonable conditions for STS measurements.



**Figure 5.** Sample  $dI/dV$  spectra in the test set that are classified as (a) *false positives* (labeled as *false*, predicted as *true*) and (b) *false negatives* (labeled as *true*, predicted as *false*) by an AdaBoost model.

Since CatBoost models are only compatible with 64-bit systems and deep neural networks require more complex environments to deploy (with only a marginal boost on performances), we recommend the application of an AdaBoost model and will use this as the default model for our tip conditioning program. We have deployed the program on a Scienta Omicron STM (Matrix version 3.2 and 3.4) and repeated tests showed that our tip conditioning software requires on average 2 to 5 h before the program classifies the tip as appropriate for STS and exits the loop. Since the program will only stop after the tip has been classified as good in two consecutive tip conditioning cycles (four  $dI/dV$  spectra recorded as discussed above), the state of the tip tends to

be stable and will not change drastically following further manipulation. We experience that our tip conditioning program can make efficient use of the idle time of the STM (e.g. unused instrument time over night or between shifts) and significantly reduces the amount of regular operating hours expended on manual tip conditioning for STS measurements.

**Possible improvements for machine learning models.** With our current machine learning models, the program may exit the tip preparation cycle at a  $dI/dV$  spectrum that features a clear gold surface state but is not entirely featureless in the region between  $-1.5$  V to  $-0.5$  V or  $+0.5$  V to  $2.0$  V region (e.g. gray shading in Figure 5a–b). In order to obtain machine learning models that can differentiate an ideal  $dI/dV$  spectra from slightly inferior ones, we require a larger library of grade 3 and 4 spectra that can be split into valid training and test sets. As the pace of expansion of our library of Au(111) surface spectra is gradual, we herein used a method to generate artificial spectra to fill this intermediate gap. In an effort to accelerate the process we set out to generate a training set of artificial spectra based on 90 grade 3 and 4 spectra identified in our library (Table 1). We used linear combinations of 25 randomly selected good spectra to generate a large library of 1000 artificial  $dI/dV$  point spectra. The randomly generated weights of the 25 spectra are *Dirichlet* distributed and sum to 1. We rated these 1000 artificial spectra as grade 3 or higher and used the library of real spectra with grade 0, 1, and 2 to train deep neural networks (training/test split ratio 80/20). While the precision scores of this machine learning algorithm exceed 97% with recall scores greater than 99% on the test set for both MLP and CNN models, when applied to 90 real grade 3 or higher spectra, only 20 spectra of this subset are classified correctly. This indicates that in this case deep neural networks do classification based on the artificial features introduced during the creation of the spectra. Regularization methods can be used to avoid overfitting so that

60 out of 90 real spectra can be classified as good, but the precision score will decrease to 70% and the number of *false positives* increases.

It is worth considering that were we to obtain a model that can differentiate grade 3 and 4 spectra from grade 0, 1, and 2, and then apply that model to our tip conditioning program, the much more stringent criteria lead to a significant increase in run time before the program successfully exits the cycle. Our current machine learning algorithms represent a practical compromise between effective use of idle instrument time and the quality factor of the SPM tip obtained through the tip conditioning process. The performance of our AdaBoost model represents a suitable balance between these two requirements and has been shown to be the most useful for the purpose of automated STM tip conditioning.

We present an automated tip conditioning program for STS measurements based on Python and machine learning. We developed an algorithm to process and analyze topographic STM images to find suitable positions for tip conditioning on clean or sparsely covered gold surfaces. For each tip conditioning cycle, we collect two consecutive  $dI/dV$  spectra at an autonomously selected tip conditioning position and use a machine learning model to determine if the spectra are commensurate with the expected  $dI/dV$  spectrum for the Au(111) surface state. If the condition of the tip is inadequate the program will instruct the STM control software to perform a poke at the same location and move to a new location before entering the next cycle. Machine learning models were trained on a library of archived  $dI/dV$  spectra on Au. Decision tree based ensemble and boosting models and deep neural networks show similar performances when identifying suitable STM tips featuring  $dI/dV$  spectra with clear gold surface state. Among these machine learning models AdaBoost was selected as the default for the automated tip conditioning software as it is robust, adaptable, and faster than other models. Our program has been implemented with a Scienta

Omicron STM but the method to process topographic images and the machine learning models can be easily transplanted to other STM platforms or software packages. We expect that our program can make efficient use of the idle time of the STM and greatly reduce the amount of valuable instrument time expended on tip conditioning for STS measurements.

**Experimental Methods.** All topographic images and  $dI/dV$  spectra were collected using a Scienta Omicron STM under ultrahigh vacuum at 4.2 K. The tips were either electrochemically etched from polycrystalline tungsten wire or manually cut from Pt/Ir (80/20) wire using a wire cutter. No significant difference in the quality of images or spectra were observed between the two types of tips. Tips were manually conditioned for imaging and then conditioned for STS measurements using the autonomous tip conditioning program.

Images and spectra presented in this paper were collected on Au(111) on mica substrates. Images and spectra were collected at a tunneling set point of  $V_{\text{bias}} = 0.05$  V,  $I_t = 20$  pA. Spectra were collected using a lock-in amplifier (Stanford Research Systems Inc., Model SR830) with a modulation frequency of 455 Hz and a modulation amplitude of 10 mV.

The operations of STM were controlled by an Omicron Matrix console and accessed by Python through the RemoteAccess\_API provided by Omicron. Machine learning models were implemented using the Scikit-Learn (0.22.1) module of Python and Keras (2.2.4) with TensorFlow backend.

## ASSOCIATED CONTENT

### Supporting Information.

The supporting information is available free of charge via the internet at <http://pubs.acs.org>.

Details of machine learning models and the correlation method.

The source code will be made available at [https://github.com/Kaiden713/auto\\_stm](https://github.com/Kaiden713/auto_stm)

## AUTHOR INFORMATION

### **Author contributions**

Ideas and tip conditioning procedures were developed by all authors. Archived  $dI/dV$  spectra were labeled by R.B. and machine learning models were developed by S.W. and J.Z. to avoid over tuning the models based on human perception.

### **Notes**

The authors declare no competing financial interest. A patent has been filed on the subject and all authors are affiliated with University of California, Berkeley.

### **ORCID**

Shenkai Wang: 0000-0002-2151-8664

Junmian Zhu: 0000-0002-3869-5974

Raymond Blackwell: 0000-0002-3501-9444

Felix R. Fischer: 0000-0003-4723-3111

## ACKNOWLEDGMENTS

Research supported by the National Science Foundation Grant No. 1807474 (software development and STM integration) and the Office of Naval Research Grant No. N00014-19-1-2503 (acquisition of STS spectra). R.B. acknowledges support through a National Science

Foundation Graduate Research Fellowship under grant DGE-1106400. The authors like to thank Prof. Michael F. Crommie, Dr. Peter H. Jacobse, and Mr. Gregg Hauser for helpful discussions and suggestions.

## REFERENCES

- (1) Feenstra, R. M.; Stroscio, J. A.; Fein, A. P. Tunneling Spectroscopy of the Si(111)2×1 Surface. *Surf. Sci.* **1987**, *181*, 295–306.
- (2) Hamers, R. J. Atomic-Resolution Surface Spectroscopy with the Scanning Tunneling Microscope. *Annu. Rev. Phys. Chem.* **1989**, *40*, 531–559.
- (3) Crommie, M. F.; Lutz, C. P.; Eigler, D. M. Confinement of Electrons to Quantum Corrals on a Metal Surface. *Science* **1993**, *262*, 218–220.
- (4) Feenstra, R. M. Scanning tunneling spectroscopy. *Surf. Sci.* **1994**, *299–300*, 965–979.
- (5) Jamneala, T.; Madhavan, V.; Chen, W.; Crommie, M. F. Scanning tunneling spectroscopy of transition-metal impurities at the surface of gold. *Phys. Rev. B* **2000**, *61*, 9990–9993.
- (6) Brar, V. W.; Zhang, Y.; Yayon, Y.; Ohta, T.; McChesney, J. L.; Bostwick, A.; Rotenberg, E.; Horn, K.; Crommie, M. F. Scanning tunneling spectroscopy of inhomogeneous electronic structure in monolayer and bilayer graphene on SiC. *Appl. Phys. Lett.* **2007**, *91*, 122102.
- (7) Chen, Y.-C.; Cao, T.; Chen, C.; Pedramrazi, Z.; Haberer, D.; de Oteyza, D. G.; Fischer, F. R.; Louie, S. G.; Crommie, M. F. Molecular bandgap engineering of bottom-up synthesized graphene nanoribbon heterojunctions. *Nat. Nanotech.* **2015**, *10*, 156–160.

- (8) Rizzo, D. J.; Veber, G.; Cao, T.; Bronner, C.; Chen, T.; Zhao, F.; Rodriguez, H.; Louie, S. G.; Crommie, M. F.; Fischer, F. R. Topological Band Engineering of Graphene Nanoribbons. *Nature* **2018**, *560*, 204–208.
- (9) Mishra, S.; Beyer, D.; Eimre, K.; Kezilebieke, S.; Berger, R.; Gröning, O.; Pignedoli, C. A.; Müllen, K.; Liljeroth, P.; Ruffieux, P.; Feng, X.; Fasel, R. Topological frustration induces unconventional magnetism in a nanographene. *Nat. Nanotech.* **2020**, *15*, 22–28.
- (10) Bartels, L.; Meyer, G.; Rieder, K.-H. Controlled vertical manipulation of single CO molecules with the scanning tunneling microscope: A route to chemical contrast. *Appl. Phys. Lett.* **1997**, *71*, 213–215.
- (11) Gross, L.; Moll, N.; Mohn, F.; Curioni, A.; Meyer, G.; Hanke, F.; Persson, M. High-Resolution Molecular Orbital Imaging Using a *p*-Wave STM Tip. *Phys. Rev. Lett.* **2011**, *107*, 086101.
- (12) Hapala, P.; Kichin, G.; Wagner, C.; Tautz, F. S.; Temirov, R.; Jelínek, P. Mechanism of high-resolution STM/AFM imaging with functionalized tips. *Phys. Rev. B* **2014**, *90*, 085421.
- (13) Albrechtsen, O.; Salemink, H. W. M.; Moench, K. A.; Thölen, A. R. Reliable tip preparation for high-resolution scanning tunneling microscopy. *J. Vac. Sci. Technol., B* **1994**, *12*, 3187–3190.
- (14) Klein, M.; Schwitzgebel, G. An improved lamellae drop-off technique for sharp tip preparation in scanning tunneling microscopy. *Rev. Sci. Instrum.* **1997**, *68*, 3099–3103.

- (15) Ernst, S.; Wirth, S.; Rams, M.; Dolocan, V.; Steglich, F. Tip preparation for usage in an ultra-low temperature UHV scanning tunneling microscope. *Sci. Technol. Adv. Mat.* **2007**, *8*, 347–351.
- (16) Salmeron, M., Scanning Tunneling Microscopy. In *Characterization of Materials*, 2002.
- (17) Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. Surface Studies by Scanning Tunneling Microscopy. *Phys. Rev. Lett.* **1982**, *49*, 57–61.
- (18) Hla, S.-W.; Braun, K.-F.; Iancu, V.; Deshpande, A. Single-Atom Extraction by Scanning Tunneling Microscope Tip Crash and Nanoscale Surface Engineering. *Nano Lett.* **2004**, *4*, 1997–2001.
- (19) Rashidi, M.; Wolkow, R. A. Autonomous Scanning Probe Microscopy in Situ Tip Conditioning through Machine Learning. *ACS Nano* **2018**, *12*, 5185–5189.
- (20) Chen, W.; Madhavan, V.; Jamneala, T.; Crommie, M. F. Scanning Tunneling Microscopy Observation of an Electronic Superlattice at the Surface of Clean Gold. *Phys. Rev. Lett.* **1998**, *80*, 1469–1472.
- (21) Freund, Y.; Schapire, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
- (22) Dorogush, A. V.; Ershov, V.; Gulin, A. CatBoost: gradient boosting with categorical features support. *arXiv:1810.11363* **2018**.
- (23) Woolley, R. A. J.; Stirling, J.; Radocea, A.; Krasnogor, N.; Moriarty, P. Automated probe microscopy via evolutionary optimization at the atomic scale. *Appl. Phys. Lett.* **2011**, *98*, 253104.

- (24) Stirling, J.; Woolley, R. A. J.; Moriarty, P. Scanning probe image wizard: A toolbox for automated scanning probe microscopy data analysis. *Rev. Sci. Instrum.* **2013**, *84*, 113701.
- (25) Møller, M.; Jarvis, S. P.; Guérinet, L.; Sharp, P.; Woolley, R.; Rahe, P.; Moriarty, P. Automated extraction of single H atoms with STM: tip state dependency. *Nanotechnology* **2017**, *28*, 075302.
- (26) Ziatdinov, M.; Maksov, A.; Kalinin, S. V. Learning surface molecular structures via machine vision. *NPJ Comput. Mater.* **2017**, *3*, 31.
- (27) Huang, B.; Li, Z.; Li, J. An artificial intelligence atomic force microscope enabled by machine learning. *Nanoscale* **2018**, *10*, 21320–21326.
- (28) Gordon, O. M.; Junqueira, F. L. Q.; Moriarty, P. J. Embedding human heuristics in machine-learning-enabled probe microscopy. *Machine Learning: Science and Technology* **2020**, *1*, 015001.
- (29) Gordon, O. M.; Moriarty, P. J. Machine learning at the (sub)atomic scale: next generation scanning probe microscopy. *Machine Learning: Science and Technology* **2020**, *1*, 023001.
- (30) Krull, A.; Hirsch, P.; Rother, C.; Schiffrin, A.; Krull, C. Artificial-intelligence-driven scanning probe microscopy. *Communications Physics* **2020**, *3*, 54.
- (31) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.