Supplementary Material for Autonomous Scanning Probe Microscopy Investigations over WS_2 and Au{111}

John C. Thomas[1*], Antonio Rossi[1,2], Darian Smalley[3], Luca Francaviglia[1], Zhuohang Yu[4,5], Tianyi Zhang[4,5], Shalini Kumari[4,5], Joshua A. Robinson[4,5], Mauricio Terrones[4,5,6], Masahiro Ishigami[3], Eli Rotenberg[2], Edward S. Barnard[1], Archana Raja[1], Ed Wong[1], D. Frank Ogletree[1], Marcus M. Noack[7*], and Alexander Weber-Bargioni[1*]

[1]*Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America*
[2]*Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America*
[3]*Department of Physics, University of Central Florida, Orlando, FL 32816, United States of America*
[4]*Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA 16082 United States of America*
[5]*Center for Two-Dimensional and Layered Materials, The Pennsylvania State University, University Park, PA, 16802 United States of America*
[6]*Department of Physics and Department of Chemistry, The Pennsylvania State University, University Park, PA, 16802 United States of America*
[7]*Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America*
[*]*jthomas@lbl.gov, marcusnoack@lbl.gov, afweber-bargioni@lbl.gov*

SUPPLEMENTARY NOTES

1| Gaussian Process Regression and Acquisition Function

The GP model can be characterized for a given dataset, $D = \{\mathbf{x}_i, y_i\}$, as the probability distribution over functions $f(x)$ as

$$p(\mathbf{f}) = -\frac{1}{\sqrt{(2\pi)^{dim}|\mathbf{K}|}} exp\left[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right] \tag{1}$$

where, by applying the covariance kernel, K is the calculated covariance matrix of D, and $\boldsymbol{\mu}$ is the prior mean vector. The likelihood over $\mathbf{y}$ is then

$$p(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^{dim}|\mathbf{V}|}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{f})^T \mathbf{V}^{-1}(\mathbf{y}-\mathbf{f})\right] \tag{2}$$

where $\mathbf{V}$ is the matrix of the possibly non-i.i.d. noise (heteroscedastic). From the above equations, the posterior probability distribution for a measurement outcome at $\mathbf{x}_0$ can be calculated. The probability distribution mean and variance functions are then defined as

$$m(\mathbf{x}_0) = \boldsymbol{\mu} + \mathbf{k}^T(\mathbf{K}+\mathbf{V})^{-1}(\mathbf{y}-\boldsymbol{\mu}) \tag{3}$$

$$\sigma^2(\mathbf{x}_0) = \mathbf{k}(\mathbf{x}_0,\mathbf{x}_0) - \mathbf{k}^T(\mathbf{K}+\mathbf{V})^{-1}\mathbf{k} \tag{4}$$

Covariances are computed by the Matérn kernel, which is commonly chosen to match physical processes.[1] The kernel is defined as

$$k(\mathbf{x}_i,\mathbf{x}_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}}{l}r\right)^{\nu} B_\nu\left(\frac{\sqrt{2\nu}}{l}r\right) \tag{5}$$

where $r = \parallel \mathbf{x}_i - \mathbf{x}_j \parallel_{l_2}$, B_ν is the modified Bessel function, and ν is a parameter defining the differentiability. This is combined with an anisotropic kernel definition to control the level of differentiability in each direction of the input space.

The GP model prior and posterior can be used in the definition of the acquisition function.[2] The acquisition function can take a variety of forms that include measuring the total uncertainty or favor morphological features. The implemented software can include a user-defined acquisition function that passes through optimization and gives the next point of measurement for data collection. Using the GP defined joint prior given as

$$p(\mathbf{f}, \mathbf{f}_0) = \frac{1}{\sqrt{(2\pi)^{dim}|\Sigma|}} \exp\left[-\frac{1}{2}\begin{pmatrix}\mathbf{f}-\mu \\ \mathbf{f}_0-\mu_0\end{pmatrix}^T \Sigma^{-1}\begin{pmatrix}\mathbf{f}-\mu \\ \mathbf{f}_0-\mu_0\end{pmatrix}\right] \tag{6}$$

where

$$\Sigma = \begin{pmatrix} K & \kappa \\ \kappa^T & \mathscr{K} \end{pmatrix} \tag{7}$$

with $\kappa_i = k(\phi, x_0, x_i)$, $\mathscr{K} = k(\phi, x_0, x_0)$, $K_{ij} = k(\phi, x_i, x_j)$, and x_0 is the point of interest. The acquisition function can now be defined as

$$f_{acq}(x) = f_{acq}\left(m(x), \sigma^2(x), \mathscr{K}, \Sigma(x)\right) \tag{8}$$

and directly provide where the next measurement point will be taken during an autonomous experiment. The simplest example of the acquisition function is the exploration method given as

$$f_{acq}(x) = \sigma^2(x) \tag{9}$$

where the posterior variance maximum is the next point of acquisition. The general definition of the acquisition function also allows for Bayesian optimization such as using the upper confidence bound (UCB) method given as

$$f_{acq}(x) = m(x) + c\sqrt{\sigma^2(x)} \tag{10}$$

and information theory methods such as the Shannon-information gain defined by

$$f_{acq}(x) = Entropy(\mathscr{K}) - Entropy(\Sigma(x)) \tag{11}$$

which can be further fine-tuned to find regions of interest, e.g. using a reference spectrum or image feature. A cost function can also be included, which defines the cost of a given measurement itself and the cost of moving within a defined parameter space.

Here we compare results from using the diagonal of the variance-covariance matrix, UCB, and using the Shannon-information gain (SIG). Additionally, we also compare results from a dense 128×128 grid, an equally spaced 12×12 grid, and purely randomly derived locations.

2| Drift Tracking

Drift within scanning probe microscopy techniques can be solved through a variety of methods either during the experiment or data manipulation post acquisition. Piezocreep and thermal drift are the two largest contributors to drift, where piezocreep is a nonlinear effect that can be minimized through correcting for the desired tip position to the actual tip location. However, as the piezos move or scan, some amount of thermal drift is expected. A number of image analysis methodologies that make use of features within a raster-scan image have shown efficacy in tracking drift between multiple images.[3] Atom-tracking, where the STM tip is locked onto an atom or molecule using two-dimensional lateral feedback can maintain spatial coordinates and be used to track drift velocity.[4] Another method that takes advantage of matching sets of key points across two images enables correction of nonlinearities within piezo positioners.[5] The concept of using image pairs is used across scanning probe microscopy and is also applicable to scanning transmission electron microscopy, where drift distortion can be corrected line-by-line by rectifying scan line origins to minimize differences between measured line intensity and the image recorded along an orthogonal direction.[6] Here, we apply one methodology towards tracking drift during point scanning tunneling spectroscopy acquisition.

3| Convolutional Neural Network Performance

Metrics from our proposed model are shown below, where we look at receiver operator characteristics to determine model performance between positive and negative classes, the confusion matrix to view sensitivity, specificity, the false negative rate, and the false positive rate, and the precision recall curve that depicts the true positive rate and the positive predictive value. In all three metrics, the chosen model performs optimally on the data provided against four classes of spectroscopic data. We compare two models with matching hyperparameters, as detailed in https://github.com/jthomas03/gpSTS, that only differ in the number of epochs that the neural network is trained. At 6 epochs, off-diagonal elements, which represent a type i (false positive) or type ii (false negative) error, within the confusion matrix are minimized.

4| gpSTS Experimental Workflow

A GP experiment can be started over a given region with some initialization parameters, which is provided by the user in the form of a n$\times$n image containing information for pixel size, experimental boundary conditions, and location. Additional information, such as file locations, experiment name, imaging parameters for drift tracking, voltage range and stepsize, and any high-level neural network hyperparameters are input into the configuration file in gpSTS. An example file is shown below.

```
nanonis_config = {
    "Nanonis_Settings": {
        "File": "gpSTSinit",
        "ExperimentName": "Test Out",
        "Version": "0.0.1",
        "ImageStart": "test_img001.sxm",
        "FolderLocation": "\\gpSTS\\src\\",
        "DataLocation": "\\gpSTS\\src\\data\\",
        "Channel": "Z",
        "ImDirection": "forward",
        "SpectralRange": [-1,1],
        "NumSpectralPoints": 1200,
        "Center_Point": [174,34],
        "Search_Window": 40,
        "Feature_Window": 20,
        "ScanCurrent": 30e-12,
        "SpecCurrent": 200e-12,
        "STSbias": "Bias calc (V)",
        "STSsignal": "Current (A)"
    },
    "Neural_Network": {
        "TrainingPath": "\\gpSTS\\src\\train\\",
        "EpochNumber": 2,
        "ClassNumber": 4,
        "LearningRate": 0.001,
        "BatchSizeTrain": 5,
        "BatchSizeVal": 1,
```

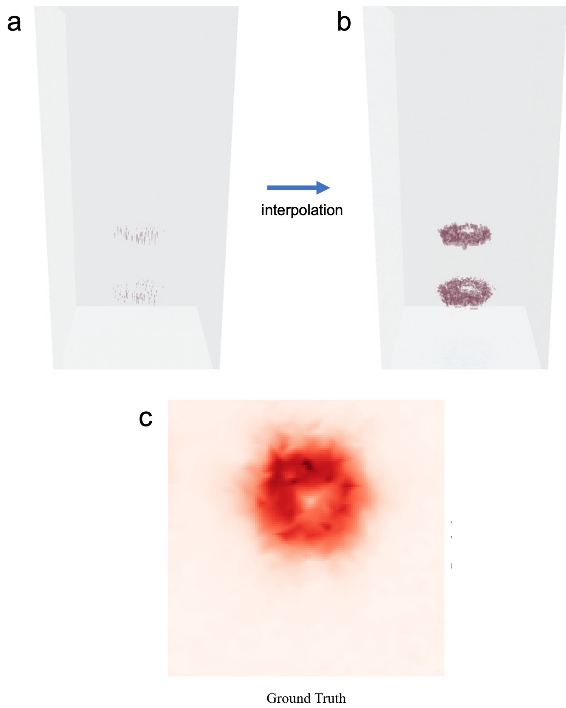
```
        "BatchSizeTest": 1
    }
}
```

Acquisition functions choice can also be modified in the configuration file and customized by the user. For more advanced customization, please refer to the documentation for the library for autonomous that is used in gpSTS, gpCAM.[2]

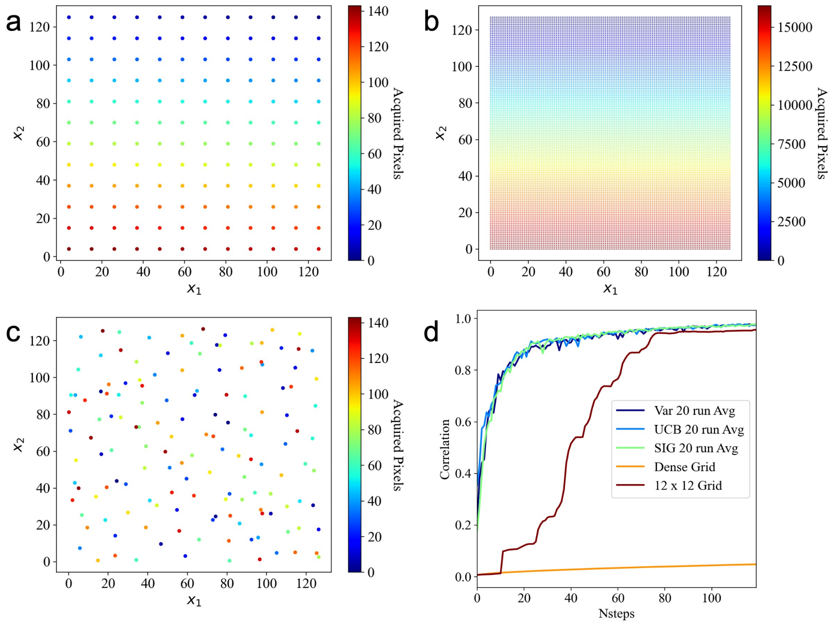
```
###acquisition functions###
def my_ac_func(x,obj):
  mean = obj.posterior_mean(x)["f(x)"]
  cov  = obj.posterior_covariance(x)["v(x)"]
  sig = obj.shannon_information_gain(x)["sig"]
  ucb = mean + 3.0 * np.sqrt(cov)
  return cov
```

The experiment is then run via command line interface with the Nanonis controller running with the option presented in the article shown below (using the signal summation in acquired spectra).

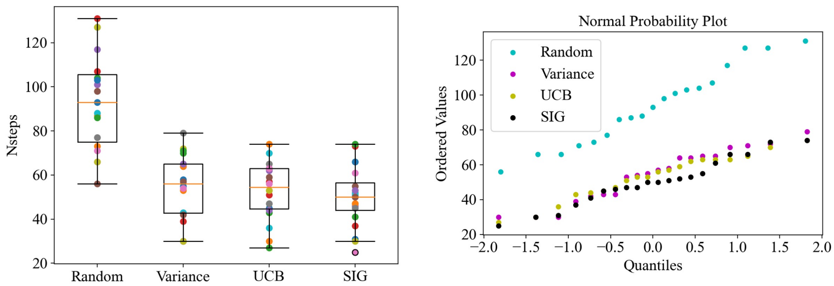
SUPPLEMENTARY FIGURES



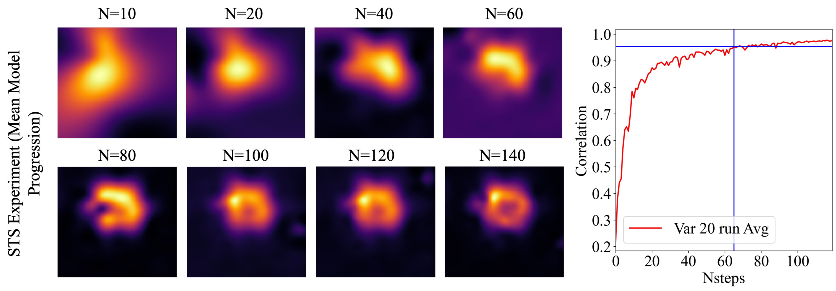
Supplementary Figure 1: **Extended Gaussian-process-driven experiment.** a) Ground truth data was formulated from a live experimental run with collected points shown in $x_1 \times x_2 \times bias_{sample}$ (V) space with spectral intensity $(dI/dV\ a.u.)$ depicted in a white to red colorscale, and then b) interpolating the data to a dense 128 x 128 datacube. c) Each point can be summed within a given range (0 to 0.7 V to capture in-gap states), which is shown as an image, and can then be used to compare different acquisition function performance while GP takes point STS shown in b and performs an equivalent summation to reach a predicted mean value function.



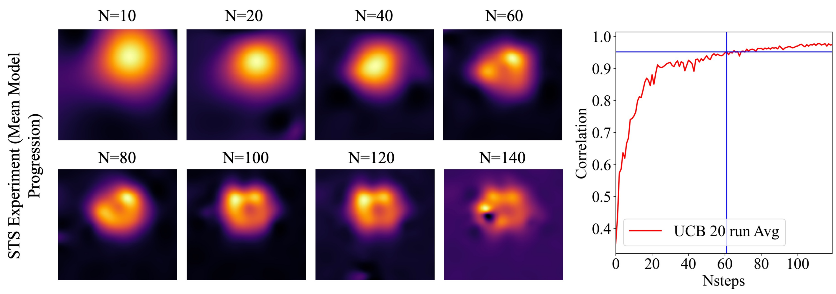
Supplementary Figure 2: **Comparison of point collection methods.** In order to test the performance of a GP on low temperature STS data, we input points along two types of grids, compute summation values, interpolate with points acquired, and fill the remaining values with prior STS minima. We show an a) equally spaced 12 x 12 grid, b) a dense 128 x 128 grid, and c) an example GP-driven grid in addition to d) correlations with ground truth data using variance, UCB, Shannon information gain, and each style of grid collection. GP-driven collection shows a measurable boost in correlation and typically reaches greater than 90 percent in under 30 points.



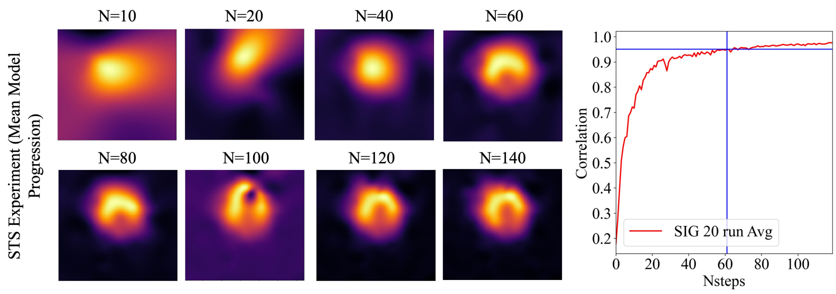
Supplementary Figure 3: **Statistical analysis for randomly-driven experimental comparison.** One-way ANOVA depicting number of iterations required to reach 95 percent correlation between a randomized experiment, variance, UCB, and SIG. Here the null hypothesis that mean values are equal is rejected (p_{value} = 3.5 x 10^{-13}). Comparing across acquisition functions alone, we fail to reject the null hypothesis over 20 runs (p_{value} = 0.61).



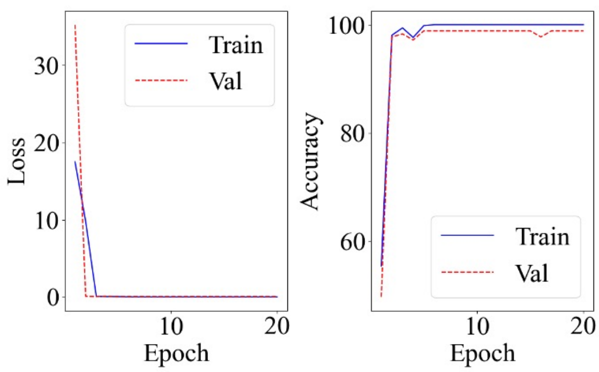
Supplementary Figure 4: **Variance.** Correlation reaches 0.954 at 65 intervals with a standard deviation of 0.00987 maintained there after and a maximum correlation of 0.981 reached.



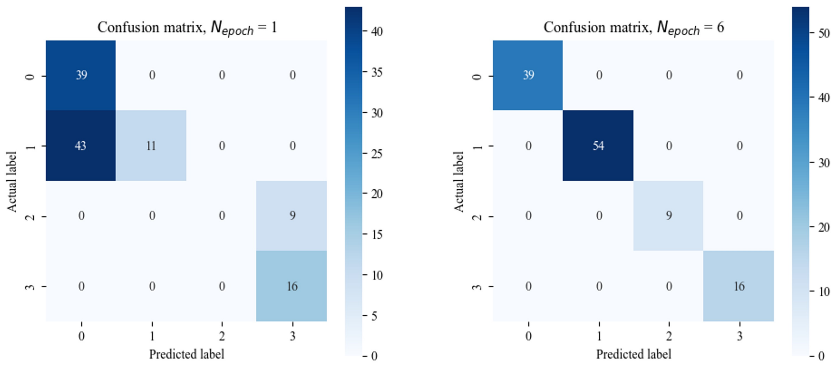
Supplementary Figure 5: **UCB.** 0.952 correlation is reached after 61 iterations with a standard deviation of 0.00999 maintained after. A maximum correlation of 0.983 is reached.



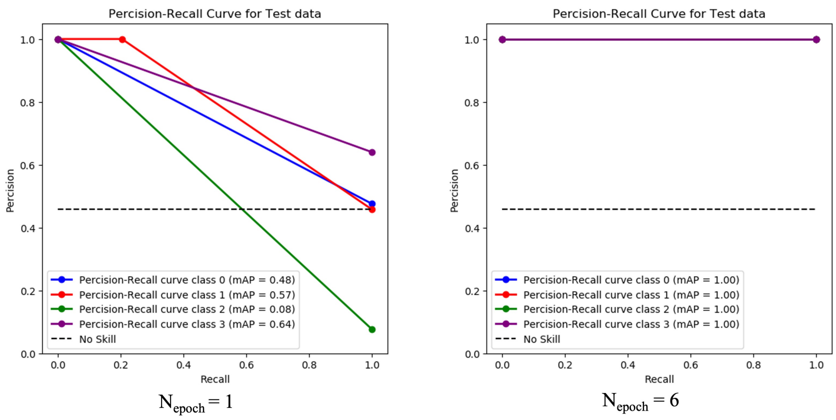
Supplementary Figure 6: **Shannon information gain.** 0.951 correlation is obtained at 61 iterations with a standard deviation of 0.00898 maintained after and a maximum correlation of 0.982 reached.



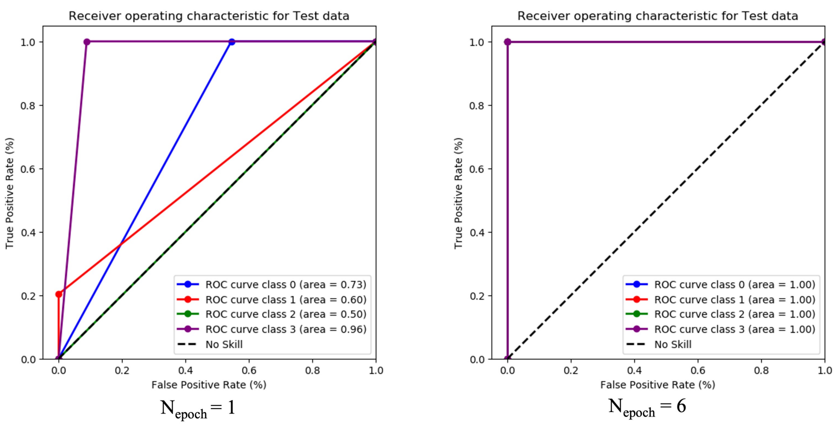
Supplementary Figure 7: **1D-CNN model performance.** Accuracy and loss after 20 training epochs is shown for completeness on both training and validation datasets.



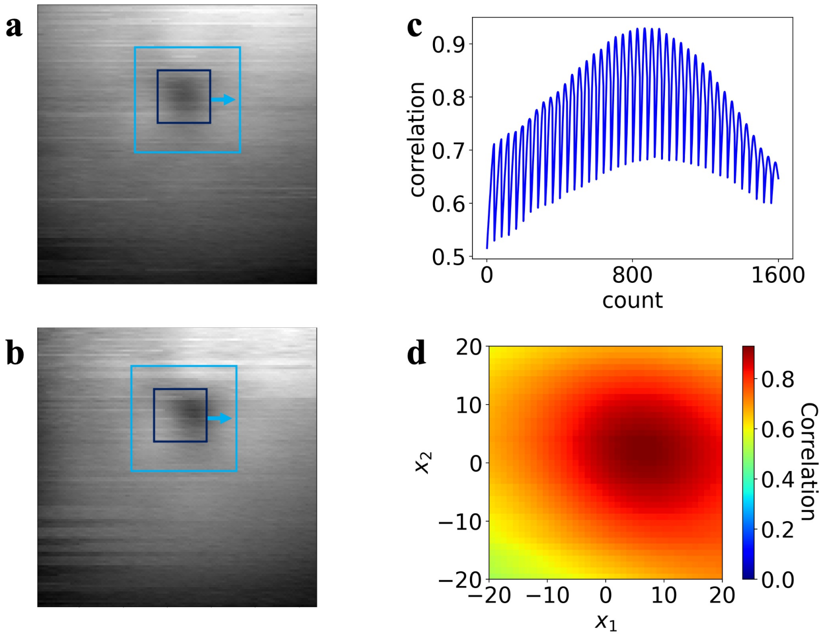
Supplementary Figure 8: **Test data classifier error summary.** Confusion matrix taken for classification using the argmax value across probabilities, where error is reduced at 6 epochs.



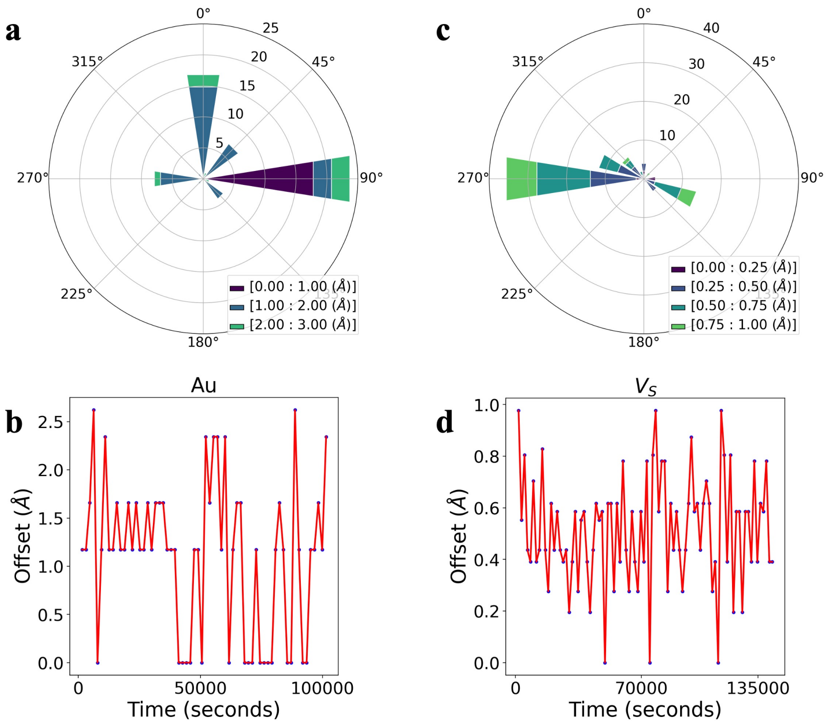
Supplementary Figure 9: **Test data precision recall curve.** Shown for test data, where the mean average precision for all classes (class 0 = Au_{fcc}, class 1 = Au_{hcp}, class 3 = V_S, class 4 = WS_2) reaches 1.0 at the end of training.



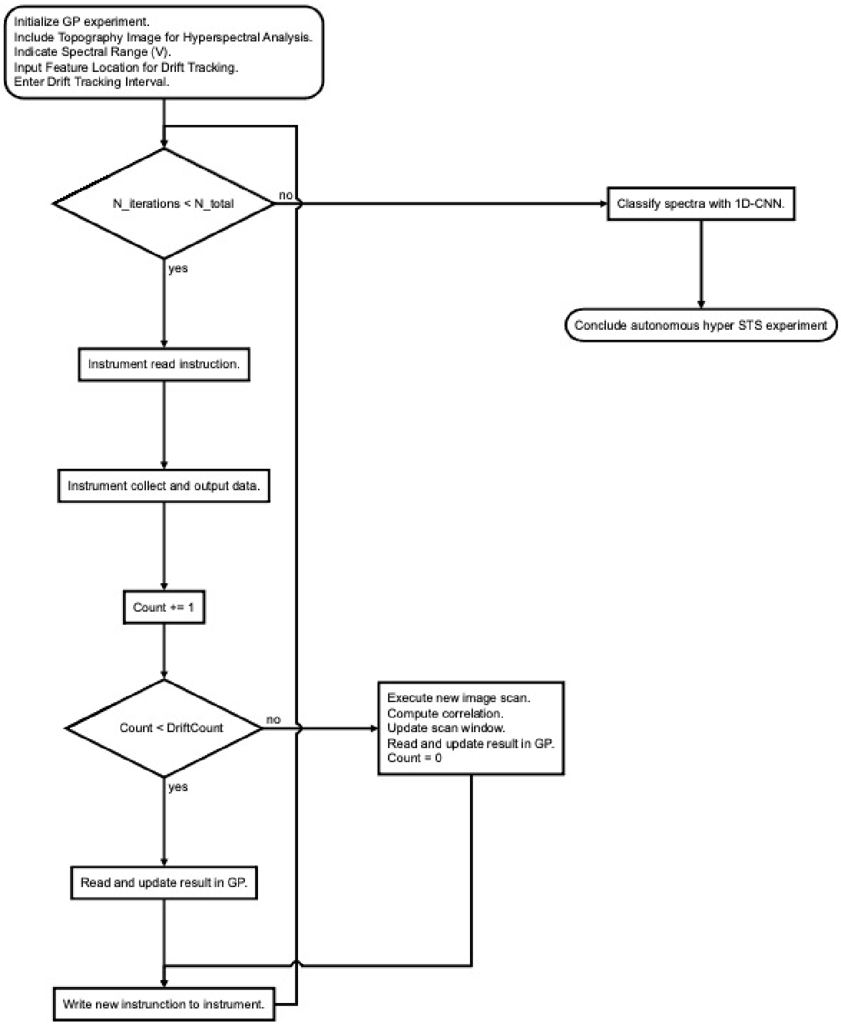
Supplementary Figure 10: **Test data ROC curve.** Shown for all classes, where AUC values indicate the model's capability to distinguish between classes after the first and last epoch used in the chosen model.



Supplementary Figure 11: **Feature correlation.** Cross-correlation over a sequential image taken at a) T = 0 and b) T = 30 min, where correlation is shown in both c) 1D and d) 2D, for clarity. Maximum correlation is obtained at $x_{pixel} = 5$ and $y_{pixel} = 1$, providing the offset in both the x- and y- directions between collected images.



Supplementary Figure 12: **Experimental drift.** Computed offsets during a Au{111} autonomous experiment is shown a) in two dimensions and b) the magnitude of the 2D vector as a function of time. Both the c) 2D and d) 1D picture are also shown for a WS_2 experiment. Drift in each example is multi-directional and highlights the need for a means to correct for any offset during measurements, however, parameters can be optimized for any possible tool-tool or run-run delta within the provided software.



Supplementary Figure 13: **Experimental gpSTS flowchart.** Workflow for gpSTS experiment where the signal summation over a given voltage range is used to run the GP, and classification is performed after spectra are acquired. STS are collected in an iterative fashion, and escapes once the number of optimal measurements are obtained.

Supplementary References

[1] Williams, C. K. & Rasmussen, C. E. *Gaussian processes for machine learning.* (MIT Press, 2006).

[2] Noack, M. M. et al. Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities. *Nat. Rev. Phys.* **3** 685-697 (2021).

[3] Mantooth, B. A. et al. Analyzing the motion of benzene on Au111: single molecule statistics from scanning probe images. *J. Phys. Chem. C* **111** 6167-6182 (2007).

[4] Swartzentruber, B. S. Direct measurement of surface diffusion using atom-tracking scanning tunneling microscopy. *Phys. Rev. Lett.* **76** 459-462 (1996).

[5] Gaponenko, I. et al. Computer vision distortion correction of scanning probe microscopy images. *Sci. Rep.* **7** 669 (2017).

[6] Ophus, C., Ciston, J., & Nelson, C. T. Correcting nonlinear drift distortion of scanning probe and scanning transmission electron microscopies from image pairs with orthogonal scan directions. *Ultramicroscopy* **162** 1-9 (2016).