

Machine learning identification of impurities in the STM images

Ce Wang(王策), Haiwei Li(李海威), Zhenqi Hao(郝镇齐), Xintong Li(李昕彤), Changwei Zou(邹昌炜), Peng Cai(蔡鹏), Yayu Wang(王亚愚), Yi-Zhuang You(尤亦庄), Hui Zhai(翟荟)

Citation: Chin. Phys. B . 2020, 29(11): 116805 . **doi:** 10.1088/1674-1056/abc0d5
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Atomic-level characterization of liquid/solid interface

Jiani Hong(洪嘉妮), Ying Jiang(江颖)

Chin. Phys. B . 2020, 29(11): 116803 . **doi:** 10.1088/1674-1056/aba9d0

Epitaxial growth of antimony nanofilms on HOPG and thermal desorption to control the film thickness

Shuya Xing(邢淑雅), Le Lei(雷乐), Haoyu Dong(董皓宇), Jianfeng Guo(郭剑峰), Feiyue Cao(曹飞跃), Shangzhi Gu(顾尚志), Sabir Hussain, Fei Pang(庞斐), Wei Ji(季威), Rui Xu(许瑞), Zhihai Cheng(程志海)

Chin. Phys. B . 2020, 29(9): 096801 . **doi:** 10.1088/1674-1056/aba27c

STM study of selenium adsorption on Au(111) surface

Bin Liu(刘斌), Yuan Zhuang(庄源), Yande Que(阙炎德), Chaoqiang Xu(徐超强), Xudong Xiao(肖旭东)

Chin. Phys. B . 2020, 29(5): 056801 . **doi:** 10.1088/1674-1056/ab821d

High quality $PdTe_2$ thin films grown by molecular beam epitaxy

En Li(李恩), Rui-Zi Zhang(张瑞梓), Hang Li(李航), Chen Liu(刘晨), Geng Li(李更), Jia-Ou Wang(王嘉鸥), Tian Qian(钱天), Hong Ding(丁洪), Yu-Yang Zhang(张余洋), Shi-Xuan Du(杜世萱), Xiao Lin(林晓), Hong-Jun Gao(高鸿钧)

Chin. Phys. B . 2018, 27(8): 086804 . **doi:** 10.1088/1674-1056/27/8/086804

Direct measurements of conductivity and mobility in millimeter-sized single-crystalline graphene via van der Pauw geometry

Rui-Song Ma(马瑞松), Qing Huan(郇庆), Liang-Mei Wu(吴良妹), Jia-Hao Yan(严佳浩), Yu-Yang Zhang(张余洋), Li-Hong Bao(鲍丽宏), Yun-Qi Liu(刘云圻), Shi-Xuan Du(杜世萱), Hong-Jun Gao(高鸿钧)

Chin. Phys. B . 2017, 26(6): 066801 . **doi:** 10.1088/1674-1056/26/6/066801

SPECIAL TOPIC — Machine learning in condensed matter physics

Machine learning identification of impurities in the STM images*

Ce Wang(王策)[1], Haiwei Li(李海威)[2], Zhenqi Hao(郝镇齐)[2], Xintong Li(李昕彤)[2], Changwei Zou(邹昌炜)[2], Peng Cai(蔡鹏)[3,†], Yayu Wang(王亚愚)[2,4], Yi-Zhuang You(尤亦庄)[5,‡], and Hui Zhai(翟荟)[1,§]

[1] *Institute for Advanced Study, Tsinghua University, Beijing* 100084, *China*

[2] *State Key Laboratory of Low Dimensional Quantum Physics, Department of Physics, Tsinghua University, Beijing* 100084, *China*

[3] *Department of Physics and Beijing Key Laboratory of Opto-electronic Functional Materials and Micro-nano Devices, Renmin University of China, Beijing* 100872, *China*

[4] *Frontier Science Center for Quantum Information, Beijing* 100084, *China*

[5] *Department of Physics, University of California, San Diego, California* 92093, *USA*

(Received 2 July 2020; revised manuscript received 7 August 2020; accepted manuscript online 14 October 2020)

We train a neural network to identify impurities in the experimental images obtained by the scanning tunneling microscope (STM) measurements. The neural network is first trained with a large number of simulated data and then the trained neural network is applied to identify a set of experimental images taken at different voltages. We use the convolutional neural network to extract features from the images and also implement the attention mechanism to capture the correlations between images taken at different voltages. We note that the simulated data can capture the universal Friedel oscillation but cannot properly describe the non-universal physics short-range physics nearby an impurity, as well as noises in the experimental data. And we emphasize that the key of this approach is to properly deal with these differences between simulated data and experimental data. Here we show that even by including uncorrelated white noises in the simulated data, the performance of the neural network on experimental data can be significantly improved. To prevent the neural network from learning unphysical short-range physics, we also develop another method to evaluate the confidence of the neural network prediction on experimental data and to add this confidence measure into the loss function. We show that adding such an extra loss function can also improve the performance on experimental data. Our research can inspire future similar applications of machine learning on experimental data analysis.

Keywords: scanning tunneling microscope, neural network, attention, data regularization

PACS: 68.37.Ef, 07.05.Mh **DOI:** 10.1088/1674-1056/abc0d5

1. Introduction

Machine learning algorithms are good at extracting features from patterns, which have found broad applications in industry such as face recognition and imaging processing. In most of physics experiments, data analysis is essentially a task of pattern recognization, and one needs to extract useful information from images taken from experiments. Therefore, recently there is a general trend of applying machine learning algorithms to data analysis in physics experiments.[1–7] The success of this approach not only can make the data analysis processes more efficiently, but also can possibly reveal hidden features and new physics behind the experimental images that are hard to be identified by human eyes.

There are different kinds applications of pattern recognization algorithms on experimental data. One typical application is clustering data with common features by unsupervised learning methods and extracting the essential common features by imposing an information bottleneck.[8] This usually works for situations where images are abundant. In this work we focus on another type of applications. In these cases, the experimental data are not abundant, but one has idea from theory about what kinds of features that we are looking for. In such cases, one can first train the neural network (NN) with big data generated by theoretical simulation of the features, and then apply the trained NN to make predictions on the experimental data. For instances, one can train an NN to learn simulated STM images with charge-density-waves of different wave vectors, and then the trained NN is applied to real STM data to recognize whether these data contain feature of any of these kinds of charge-density-waves.[3] One can also train a NN to learn the real-space single-site density distribution of the Fermi Hubbard model generated by different approximated theories, and then the trained NN is applied to quantum gas microscope data to tell which theory fits the data better.[5]

However, while we need the simulation data to train the NN, we also need to note the differences between the simu-

*HZ is supported by Beijing Outstanding Scholar Program, the National Key Research and Development Program of China (Grant No. 2016YFA0301600), and the National Natural Science Foundation of China (Grant No. 11734010). YZY is supported by a startup fund from UCSD. PC is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

†Corresponding author. E-mail: pcai@ruc.edu.cn

‡Corresponding author. E-mail: yzyou@ucsd.edu

§Corresponding author. E-mail: hzhai@tsinghua.edu.cn

© 2020 Chinese Physical Society and IOP Publishing Ltd

lation and the reality. The differences are mainly two folds. Firstly, in real systems there inevitably exist noises. In many cases, it is hard to quantify the degree of noises in reality. Secondly, most theoretical simulations can only capture the universal part of a feature and cannot capture the non-universal part. Usually, the long-range and low-energy part is universal, but the short-range and high-energy part depends on various details, which varies between different samples. We will elaborate these differences in detail in the example that we will discuss in this work. Therefore, we should prevent the NN from overfitting the simulation data, otherwise the trained model will not generalize well to realistic cases. In this work, we propose two approaches to avoid overfitting these unrealistic features in the simulation data, that are the data augmentation and the confidence regularization, respectively. One of the key points in this work is to show that, when an NN is trained with the simulated data, whether and how one mitigates the overfitting of unrealistic features can significantly change the outcome when the trained NN is applied to experimental data. Therefore, one should be very careful when using this stratagem to judge a theory.[5] We will demonstrate this point by a concrete task, that is to identify the locations of impurities from images obtained by the scanning tunneling microscope (STM).

2. Identification of impurities in STM imaging

The STM measurements result in two-dimensional maps of the local electron density-of-state (DoS) at the surface of a material. For a given surface of a sample, by varying the voltage bias, one can obtain a sequence of maps, which correspond the local DoS at different Fermi energies. Usually there are always impurities in the materials, and these impurities can manifest themselves in the local electron DoS. Our goal is to identify the locations of the impurities from these images. Therefore, we would like to build an NN to learn the mapping from the STM images $\boldsymbol{X}^l$ to the distribution of impurities Y^l, where l labels samples in the data set. More concretely, each input of the NN $\boldsymbol{X}^l$ contains a sequence of two-dimensional images as $\{X_i^l, \{i = 1, \ldots, N_v\}\}$, where i labels images on the same surface but with different Fermi energies. N_v is the total number of images measured at a given surface. Y^l is a two-dimensional probability distribution function, which is normalized to unity.

Here we should emphasize that different X_i^l with the same l are actually correlated, because if the presence of one impurity manifests itself in the locally DoS with one voltage, the similar features must consistently appear in images with other voltages. On the other hand, if similar feature occurring in one image comes from noises, it will not show up in images with other voltages. That is to say, if we properly capture the correlations between images with different voltages, it can help us detect impurities more accurately.

3. Design of the NN

To capture the correlation between images taken at different voltages, we implement attention mechanism in our NN. Attention mechanism has been widely used in nature language processing,[9,10] which can capture the long distance correlation between different sentences. Here we treat the STM images from different voltages as different "sentences", and apply the self-attention approach to aggregate the consistent features among these images. More concretely, our NN structure is shown in Fig. 1. As mentioned above, each input data $\boldsymbol{X}^l$ is made of a sequence of images X_i^l. Here we consider each X_i^l as an $L \times L$ pattern. Here we first apply the convolutional NN to general three tensors Q_i^l, K_i^l, and V_i^l. The convolutional NN is commonly used for extracting features from images. In term of the terminology of attention, they are called the query, the key, and the value, respectively. Their dimensions are $L \times L \times d_q$, $L \times L \times d_k$, and $L \times L \times d_v$, respectively. Following the scheme of attention, we introduce the normalized value $\tilde{V}_i^l$ as

$$\tilde{V}_i^l = \frac{1}{Z_i^l} \sum_j V_j^l \exp\left(\langle Q_i^l K_j^l \rangle\right), \tag{1}$$

where $\langle Q_i^l K_j^l \rangle$ denotes the inner-product between the query and the key. For taking the inner-product, we need $d_q = d_k$ and $Z_i^l = \sum_j \exp(\langle Q_i^l K_j^l \rangle)$. After the attention, we sum over all $\tilde{V}_i^l$ and obtain $V_{\mathrm{T}}^l = \sum_i \tilde{V}_i^l$. We apply a linear projection to project V_{T}^l from $L \times L \times d_v$ onto $L \times L$ dimensional $\mathcal{S}^l$. Finally, by applying a Softmax to $\mathcal{S}$, we obtain an $L \times L$ dimensional image Y_{net}^l as the output of the NN. The Softmax function automatically ensures that Y_{net}^l is normalized to unity. This output should be compared with the labelled impurity distribution

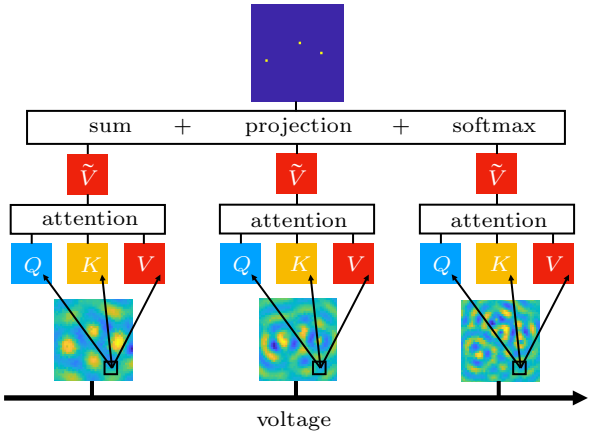


Fig. 1. The structure of the NN. The voltage arrow indicates a sequence of images taken at different voltages. Each image X_i^l is mapped to tensors of the query Q_i^l, the key K_i^l, and the value V_i^l. By the attention mechanism, it leads to normalized value $\tilde{V}_i^l$ computed by Eq. (1). Then, we sum over all $\tilde{V}_i^l$ and reach the output following a linear projection and a Softmax.

$Y^l_{\rm label}$. The loss function can be defined as the cross entropy between $Y^l_{\rm net}$ and $Y^l_{\rm label}$,

$$\mathcal{L}_0 = -\sum_l Y^l_{\rm label} \odot \ln Y^l_{\rm net}, \tag{2}$$

where the operator $\odot$ means an element-wise multiply followed by a summation over all matrix elements. Minimizing the cross entropy $\mathcal{L}_0$ aims to bring the predicted impurity distribution $Y^l_{\rm net}$ close to the ground truth $Y^l_{\rm label}$ as much as possible. The NN can be trained by the standard gradient decent method.

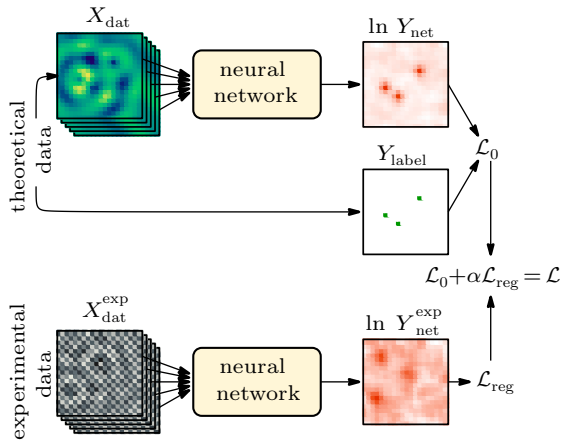


Fig. 2. Computation graph for the loss function. The NN is trained on the theoretical data via the regression loss $\mathcal{L}_0$, and regularized by the experimental data via the prediction confidence regularization $\mathcal{L}_{\rm reg}$.

4. Training data generation

Here we describe how we prepare the training data according to the theoretical knowledge about impurities. Suppose we have N_i impurities in an area discretized into $L\times L$ pixels, and their locations are $\boldsymbol{r}_w$ $(w=1,\ldots,N_i)$. The presence of an impurity induces the Friedel oscillation in a metal, and all these impurities lead to a density modulation as

$$\delta n(\boldsymbol{r}) = \sum_{w=1}^{N_i} A_w \frac{\cos(2k_{\rm f}|\boldsymbol{r}-\boldsymbol{r}_w|+\alpha_w)}{(|\boldsymbol{r}-\boldsymbol{r}_w|+\varepsilon)^2}. \tag{3}$$

We have taken the spacing between pixels as unity, such that all the other quantities are also made dimensionless accordingly. Each set of impurity position $\{\boldsymbol{r}_w, w=1,\ldots,N_i\}$ gives rise to a fixed impurity configuration, which is labelled by l. By varying the impurity position $\{\boldsymbol{r}_w\}$, we can easily generate a large number of data. We estimate the charge carrier density of the material, from which we can determine a typical value $k^0_{\rm f}$. For a given impurity configuration, we sample a number of different $k_{\rm F}$ around $k^0_{\rm F}$, which correspond to measurements at different voltages labelled by i. For a given impurity configuration and a given $k_{\rm F}$, we randomly choose parameters $\{A_w\}$ and $\{\alpha_w\}$. A_w is governed by the strength of the impurity, and α_w depends on the short-range condition of the impurity. For each $\delta n^l_i(\boldsymbol{r})$, we calculate $\bar{n}^l_i = \int \delta n^l_i(\boldsymbol{r})$, and by subtracting $\bar{n}^l_i$, we can make $\delta n^l_i(\boldsymbol{r})$ averaged to zero. Moreover, we normalize the amplitude of δn^l_i such that δn^l_i always lies between $[-1,1]$. This δn^l_x is take as input X^l_i. In this way, we generate the input $\boldsymbol{X}^l$. For the output, if there is an impurity located at one pixel, we assign this pixel a probability $1/N_i$. Otherwise, we assign it zero probability. In this way, we generate Y^l as the output.

Here we would like to highlight the difference between these generated data and the reality. Firstly, there is no noise in these data. We propose to add a white noise in each pixel which obeys a Gaussian distribution with width η. Although in reality the noises are not uncorrelated white noises and it is hard to know the actual noise distribution in reality, adding uncorrelated white noise is the most natural choice. Below we will compare the prediction power of the NN trained by data with and without noises. Secondly, we should point out that, strictly speaking, the universal form of the Friedel oscillation Eq. (3) is only valid at a distance $\gtrsim 1/k_{\rm F}$ away from an impurity. Close to an impurity, the density distribution strongly depends on the detailed potential form of the impurity. Since we do not have prior knowledge of the impurity potential, we cannot simulate accurately the short-range part of the impurity potential. In Eq. (3) we simply add a cutoff ε to avoid the divergence when $\boldsymbol{r}\to\boldsymbol{r}_w$. Therefore, the simulated data cannot describe the short-range physics nearby an impurity properly. Hence, if the NN overfits the short-range features of the training data, it can actually hinder the preference of the NN on actual experimental data. We will also discuss how to deal with this issue below.

5. Results

Here we train the NN by a large number of simulated data with $L=20$. The experimental data is taken at one sample of the gold surface, which contains a sequence of images taken at difference bias voltages. The experimental details can be found in Appendix A. We work on an area of 80×80, which is further divided to sixteen 20×20 images. We apply the NN to examine each of the 20×20 images individually. Then, we pack the predictions together and compare with a reference measurement. To properly compare different images, instead of plotting Y, we consider the layer $\mathcal{S}$ before the Softmax function. We take the local maxima of $\mathcal{S}$ as the locations of impurities. The result is shown in Fig. 3.

Here let us briefly discuss the reference measurement, and we also refer reader to Appendix A for more details. As we have discussed above, the experimental images to be recognized are taken for the local DoS around the Fermi energy. On the other hand, we can also take the experimental data at another voltage far away from the Fermi energy, which actually measures the bound state due to the impurity potential. Therefore, this measurement essentially reveals the amplitude of the

bound state wave function. We use expectation-maximization algorithm[8] to decide the positions of impurities for the bound state measurement and the results are taken as the reference answer. As shown in Fig. 3, the red squares are impurity locations predicted by the trained NN and the black points are impurity locations given by the reference measurement. Though our NN has absolutely no access to the bound state data, we can see that some predicted peaks perfectly coincide with the reference answer. We count the reference points whose distance from the nearest predicted peak is less than one pixel, which means nearly perfect agreements, and we determine the ratio of such points to the total reference points. We use this ratio to measure the preference of the NN.

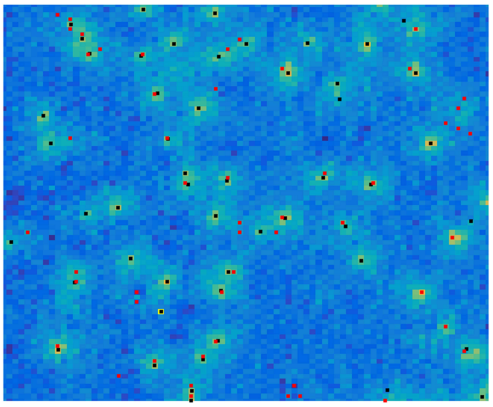


Fig. 3. Comparison between the prediction of the NN and a reference answer from another experiment. Red squares are the prediction of NN, determined by local maxima of the $\mathcal{S}$ layer generated by the NN. Black dots are the reference answer from another experiment, determined by the local maxima of the DoS of bound states shown as the color plot. Here, for training the NN, we use loss function $\mathcal{L}_0 + \alpha\mathcal{L}_{\rm reg}$ with $\alpha = 0.03$ and training data with $\eta = 0.8$ Gaussian noise.

Effect of data augmentation In Fig. 4 we plot the ratios as a function of the training epoch of the NN. We can see that the ratios increase as the training epochs increase. The blue line is given by NN trained by data without noises and the purple line is given by NN trained by data with noises. It is very clear that the performance of the NN can be significantly improved by including noises in the training data. Here we have taken $\eta = 0.8$ for noise distribution.

Effect of confidence regularization As we have mentioned above, another difference between the simulated data and the reality is the short-range physics. When the NN begins to learn the short-range details, it experiences frustration when applied to the real data, and consequently, the predictions on the real data become less confident. Here we invent a new scheme to deal with this problem. The essential idea is to force the NN to make confident predications on the real data. At each training step, we can apply the NN to the real experimental data and the NN returns an output, denoted by $Y_{\rm net}^{\exp(m)}$, where m labels different experiment regions. There could be different ways to define and evaluate the confidence level of the prediction. A natural choice is to look at the entropy of the predicted impurity distribution

$$\mathcal{L}_{\rm reg} = -\sum_m Y_{\rm net}^{\exp(m)} \odot \ln Y_{\rm net}^{\exp(m)}. \tag{4}$$

If the NN can make confident prediction on the real data, the predicted distribution will be sharply concentrated at a few positions. A sharper distribution will be more informative and hence has less entropy. We introduce $\mathcal{L}_{\rm reg}$ as an confidence regularization to our loss function, and write the full loss function $\mathcal{L}$ as $\mathcal{L}_0 + \alpha\mathcal{L}_{\rm reg}$, as illustrated in Fig. 2. When we minimize this loss function, it minimizes $\mathcal{L}_0$, which drives the NN to learn from the simulated training data, and simultaneously, it also minimizes $\mathcal{L}_{\rm reg}$, which encourages the NN to make sharp predictions on the real data. In this way, the real data is also included to regularize the training process. Here α controls the strength of the regularization. In Fig. 4, $\alpha = 0.03$ is taken to achieve the best performance. It is also clear that the performance improves significantly by including the confidence regularization. The best performance is achieved with both proper confidence regularization and adding noises in simulated data.

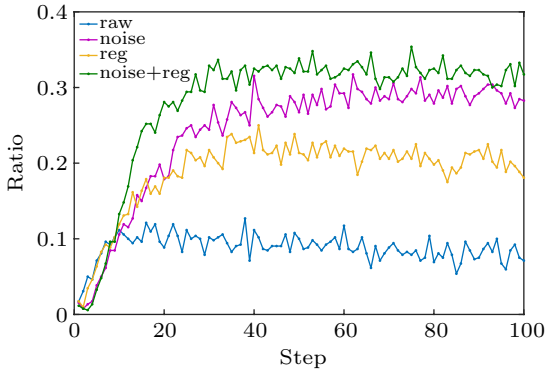


Fig. 4. Compare the performance of different approaches on the experimental data. The curves are the ratio of the reference points perfectly agreed with prediction to the total reference points. The blue line and the purple line are given by NN trained by loss function $\mathcal{L}_0$, and the training data have no noise for the blue line and include noise for the purple line. The yellow line is given by NN trained by loss function $\mathcal{L}_0 + \alpha\mathcal{L}_{\rm reg}$ including confidence regularization and data without noise. Finally, the green line is given by NN trained by loss function $\mathcal{L}_0 + \alpha\mathcal{L}_{\rm reg}$ and data with noise. Here we take $\alpha = 0.03$. The ratios are plotted as a function of the training epoch, and the results are averaged over ten independent training processes.

6. Concluding remarks

In physics research, it is quite often that we need to verify certain theory with experimental data. With the help of the machine learning algorithm, what we can do is to train the NN to learn data generated by theory, and then apply the trained NN to detect whether certain features exist in the experimental data. The success or failure in detecting this feature can be viewed as a support or a falsification of the corresponding theory. This can become a promising canonical method in future physics research. Here we point out that how to properly treat the non-universal part of the feature and the presence of

noise in reality can be crucial in this approach. The attempts of this work can be quite useful in future developments of this approach.

Appendix A: Experimental details

Here our experimental data are based on the quantum interference of Au(111) surface states. Figure A1(a) shows the typical spectra measured on the surface. A sharp increase of $\mathrm{d}I/\mathrm{d}V$ above $V = -520$ mV corresponds to the appearance of two-dimensional surface state. The dispersion of the surface band is parabolic with an effective mass ration of 0.26.[11] The data feed to NN are a series of differential conductance maps $\mathrm{d}I/\mathrm{d}V(r, eV)$ taken over 375 Å $\times$ 375 Å area, where r is the position and V is the sample bias voltage. The coherent quantum interference is dominant near Fermi level, which can be evidenced by strong spatial oscillation in $\mathrm{d}I/\mathrm{d}V$ maps. To minimize the error to the output from the data noise, our experimental dataset in this work focuses on the energies near the Fermi level, ranging from -230 meV to $+20$ meV, with equal spacing of 2 meV, as shown in Fig. A1(b). These data are used for recognition by NN. Alternatively, we turn to extract the scattering centers by imaging the defect state distribution in the $\mathrm{d}I/\mathrm{d}V$ map at $V = -520$ meV, since there are localized bound states on defects below the itinerant surface band, as shown in Fig. A1(a).

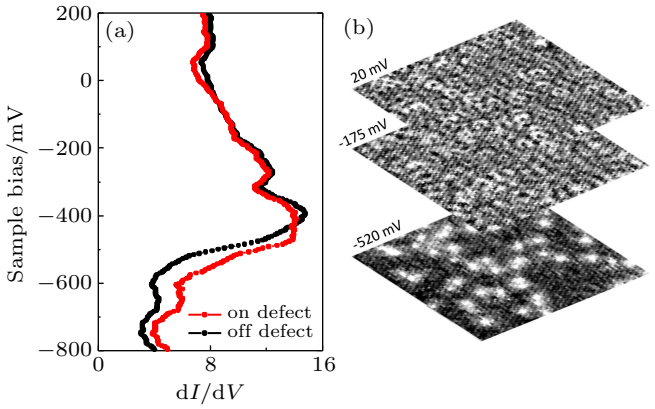


Fig. A1. Experimental data on Au(111) surface. (a) Typical $\mathrm{d}I/\mathrm{d}V$ spectra measured on defect (red curve) and on defect-free area (black curve). The abrupt increase of spectra above -520 mV shows the itinerant surface states on Au(111). Additional spectra on black curve indicates the bound state of defect below the surface band. (b) Differential conductance maps. Top panel, two representative $\mathrm{d}I/\mathrm{d}V$ maps (at -20 mV and -175 mV) taken at 5 K with tunneling junction $R = 5$ GΩ, showing spatial oscillation from the quantum interference of surface states. Bottom, the conductance map at -520 mV, revealing the spatial distribution of the defect states.

References

[1] Ziatdinov M, Maksov A, Li L, Sefat A, Maksymovych P and Kalinin S 2016 *Nanotechnology* **27** 475706

[2] Torlai G, Timar B, van Nieuwenburg E P L, Harry L, Omran A, Keesling A, Bernien H, Greiner M, Vuletić V Lukin M D, Melko R G and Endres M 2019 *Phys. Rev. Lett.* **123** 230504

[3] Zhang Y, Mesaros A, Fujita K, Edkins SD, Hamidian MH, Ch'ng K, Eisaki H, Uchida S, Davis JCS, Khatami E and Kim EA 2019 *Nature* **570** 484

[4] Rem B S, Käming N, Tarnowski M, Asteria L, Fläschner N, Becker C, Sengstock K and Weitenberg C 2019 *Nat. Phys.* **15** 917

[5] Bohrdt A, Chiu C S, Ji G, Xu M, Greif D, Greiner M, Demler E, Grusdt F and Knap M 2019 *Nat. Phys.* **15** 921

[6] Samarakoon A M, Barros K, Li Y W, Eisenbach M, Zhang Q, Ye F, Dun Z L, Zhou H, Grigera S A, Batista C D and Tennant D A 2020 *Nat. Commun.* **11** 892

[7] Khatami E, Guardado-Sanchez E, Spar B M, Carrasquilla J F, Bakr W S and Scalettar R T 2020 arXiv:2002.12310 [cond-mat.str-el]

[8] Bishop C M 2006 *Pattern recognition and machine learning* (New York: Springer) pp. 592–595

[9] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 arXiv:1706.03762 [cs.CL]

[10] Devlin J, Chang M W, Lee K and Toutanova K 2018 arXiv:1810.04805 [cs.CL]

[11] Chen W, Madhavan V, Jamneala T and Crommie M F 2018 *Phys. Rev. Lett.* **80** 1469