

STM and STS

Naoya Kawakami

April 1, 2022
version 3.0

Contents

Chapter 1	Principles of STM and STS	1
1.1	Basic principles of STM	1
1.2	Tunneling effect	2
1.3	Basic principle of STS	3
1.4	Advanced expression of tunneling in STM	4
Chapter 2	Lock-in amplifier	9
2.1	Basic principle	9
2.2	Cancellation of noise by lock-in detection	10
2.3	Setting the phase in STS measurement	10
2.4	LPF and time constant	11
2.5	Broadening of spectral feature	13
Chapter 3	Distortion of STM images	15
3.1	Thermal drift	15
3.2	Creep	16
Chapter 4	Interference of electrons on surfaces	19
4.1	Interference of one-dimensional case	19
4.2	QPI and band structure	20
4.3	Analysis of QPI	22
Chapter 5	Appendix	25
5.1	Quantum tunneling of an electron between tip and sample	25
5.2	Calculation detail to get Equation 1.19	28
5.3	Detection of second derivative in Lock-in	30
5.4	Phase sensitivity of Lock in detection	31
5.5	Calculation of thermal drift	32
5.6	Experiential error in STM observation (STM 2)	34
5.7	Affine calibration	36
Bibliography		37

Chapter 1

Principles of STM and STS

Scanning tunneling microscopy (STM) was invented by G. Binnig and H. Rohrer in 1982 [1]. With its outstanding resolution, STM has had significant impacts on the field of atomic-scale science. Especially, the atomic resolution images of Si(111)- 7×7 provided conclusive evidence to determine the complicated surface atomic structure of Si(111) 7×7 [2]. Binnig and Rohrer were awarded the Nobel prize in 1986 for the invention of STM. This chapter introduces the basic principles and characteristics of STM and scanning tunneling spectroscopy (STS).

1.1 Basic principles of STM

Figure 1.1 illustrates the basic concepts of STM. An atomically sharp tip is used as a probe to scan the surface. When the tip and sample come close to each other, a tunneling current flows. Because the value of the tunneling current strongly depends on the distance between the tip and sample, it is used as a feedback signal to keep the tip-sample distance constant. As a result, the tip path during the scanning reflects the atomic scale corrugation of the sample surface, such as atomic lattice, steps, defects, and adsorbates.

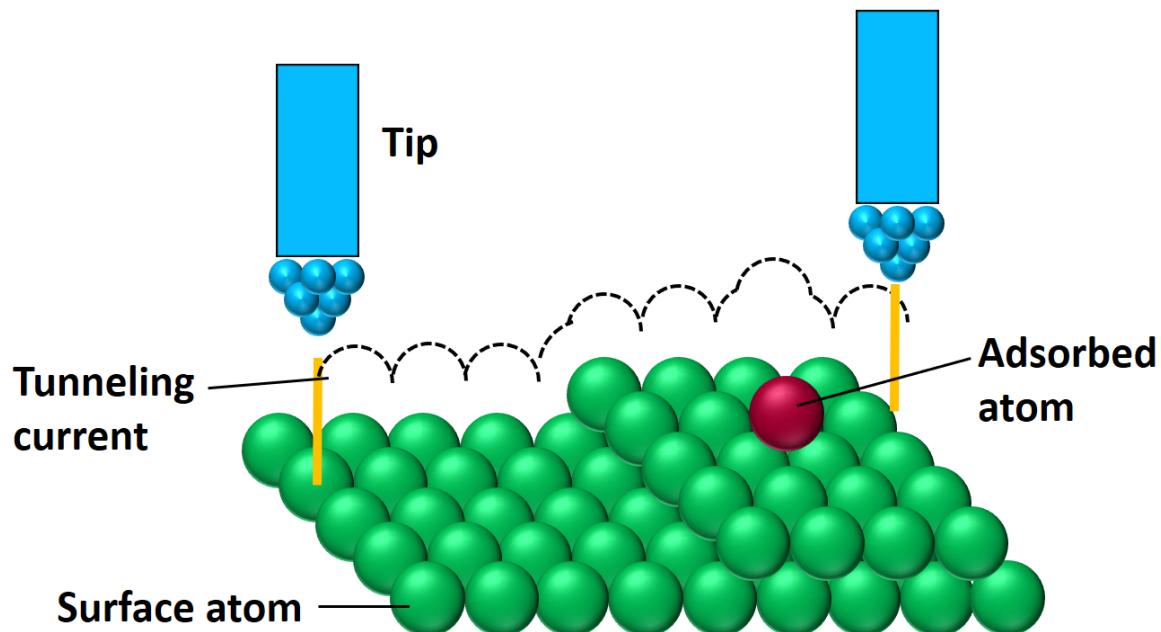


Fig. 1.1 Schmatic illustration of the basic concepts of STM

1.2 Tunneling effect

The strong distance dependence of the tunneling current is the key to realize the atomic resolution imaging of STM. With using a simple one-dimensional model, the tunneling probability of an electron between the tip and sample is expressed as follows [10] [11]^{*1}:

$$T(d, eV, E) \approx \exp \left(-\frac{2d\sqrt{2m}}{\hbar} \sqrt{\frac{\phi_s + \phi_t}{2} + \frac{eV}{2} - E} \right). \quad (1.1)$$

d: distance between the tip and sample
m: mass of an electron
 ϕ_t, ϕ_s : work function of the tip and sample
 e : elementary charge
 V : bias applied between the tip and sample
 E : energy of an electron

Figure 1.2 shows the schematic illustration of the energy diagram during STM operation. The states filled by electrons are indicated by the blue shaded part. The applied bias shifts the Fermi levels, inducing a current flow from a negatively biased electrode to the other. The number of tunneling electrons at an energy E is expressed as follows:

$$n_e(E) = \rho_t(E - eV)\rho_s(E)T(d, eV, E) \quad (1.2)$$

, which is the product of the density of states (DOS) of the tip and sample, and the tunneling probability. The net tunneling current is calculated by integrating eq. 1.2 from eV to 0. Combining with eq. 1.1, the

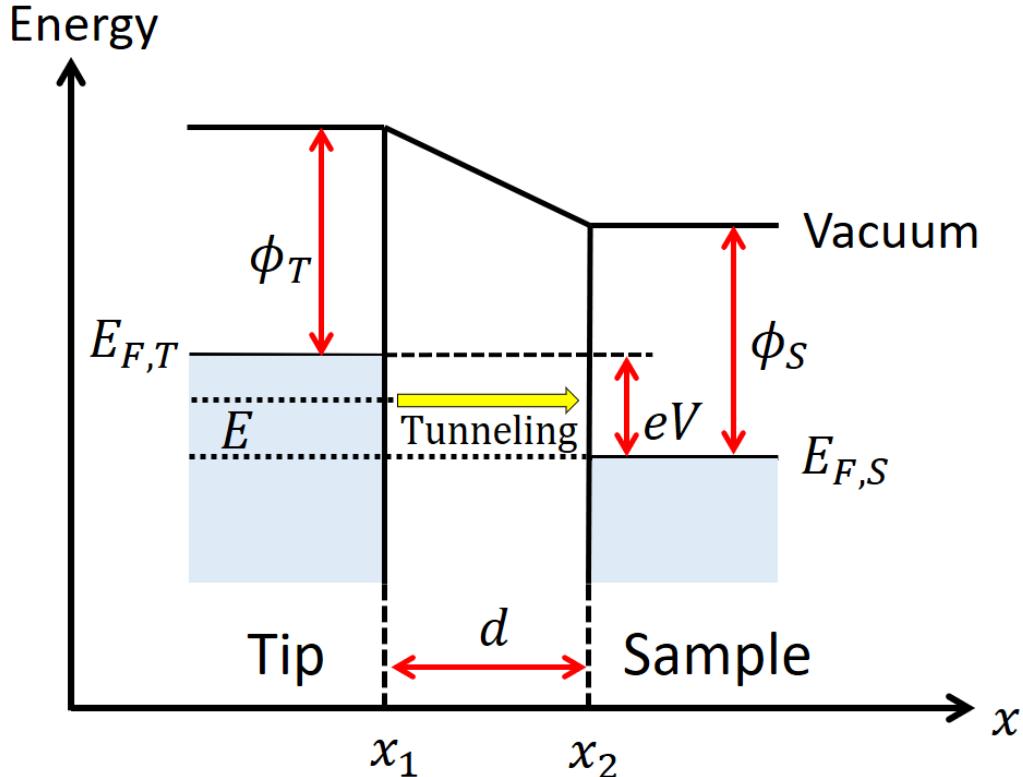


Fig. 1.2 Schmatic illustration of tunneling of electrons during the STM operation.

^{*1} see appendix if you interest

tunneling current is expressed as follows:

$$I = \int_0^{eV} \rho_t(E - eV) \rho_s(E) T(eV, E) dE \quad (1.3)$$

$$\approx I_0 \exp \left(-\frac{2d\sqrt{2m}}{\hbar} \sqrt{\frac{\phi_s + \phi_t}{2}} \right). \quad (1.4)$$

From eq. 1.3 to eq. 1.4, bias is assumed to be small enough compared with the work functions. Substitution of constants in eq. 1.4 gives a quantitative expression of tunneling current as follows: ^{*2}

$$I \approx I_0 \exp(-d \times 10^{10}). \quad (1.5)$$

Equation 1.5 shows the strong distance dependence of the tunneling current. When the distance changes from d_0 to d_1 , the change of the tunneling current can be expressed as follows:

$$\log \left(\frac{I_0}{I_1} \right) = (d_1 - d_0) \times 10^{10} = \Delta d \times 10^{10}. \quad (1.6)$$

This equation indicates that a change in distance of only 1 Å ($=1 \times 10^{-10} m$) increases or decreases the tunneling current by an order. This strong dependence of the tunneling current on the distance makes STM sensitive to the atomic scale corrugation of the sample surface.

1.3 Basic principle of STS

We have seen in eq. 1.3 that the tunneling current is a function of the DOS of the sample and tip. Therefore, the electronic structure of the sample can be deduced from the tunneling current. The relationship between the tunneling current and the DOS becomes clear by differentiating eq. 1.3 as follows [10]:

$$\begin{aligned} \frac{dI}{dV} &\propto \rho_t(0) \rho_s(eV) T(d, eV, eV) \\ &+ \int_0^{eV} \left[\rho_t(E - eV) \rho_s(E) \frac{\partial T(d, eV, E)}{\partial V} \right] dE. \end{aligned} \quad (1.7)$$

If the DOS of the tip is constant ($\rho_t = const.$), and the bias is small enough to consider the tunneling probability is constant ($T = const.$), eq. 1.7 is simplified as follows:

$$\frac{dI}{dV} \propto \rho_s(eV) \quad (1.8)$$

, which indicates that the derivative of the tunneling current is proportional to the DOS of the sample. Thus, the derivative of the I-V curve provides the bias-dependent DOS at the measurement point. This measurement is called STS. Combining with the atomic-scale sensitivity of STM, STS provides the local electronic structure of the sample.

If the bias applied between the tip and sample is large, we can not assume the tunneling probability constant, and the second term in eq 1.7 acts as a background in the spectra. For reducing the background, the dI/dV is sometimes normalized as follows: [3]:

$$\frac{dI/dV}{I/V} = \frac{d(\log I)}{d(\log V)}. \quad (1.9)$$

This is called normalized differential conductance. If the energy range of the STS measurement is large (over 1 eV), it is recommended to use the normalized value.^{*3}

^{*2} Here, 4eV is assumed as a representative value for the work functions

^{*3} Acutually I have never used this function, because I always interested in the feature around E_f . If you have some concern about the STS spectra with high bias range, consider using the normarized value.

1.4 Advanced expression of tunneling in STM

The simple form of the tunneling current is introduced in section 1.2 to show the sensitivity of STM. However, the discussion in section 1.2 is too much simplified. For the complete description of the tunneling current in quantum mechanics, the correlation with the wave function of the tip and sample should be considered. In this section, the general expression of a tunneling current is introduced according to Bardeen's theory [13]. Then the application of the theory on STM is also discussed based on the Tersoff-Hamann model [14] [15]. *⁴

1.4.1 Bardeen's transfer Hamiltonian method

A transfer matrix characterizes the electron transfer from one electrode to the other. Bardeen developed a theory to express the transfer matrix by the wave functions of two electrodes [13], which is directly applied to the STM system by considering the tip and sample as the two electrodes. Figure 1.3 shows the schematics of the system considered in this section. The potential in the tip and sample is described as U_T and U_S . When the tip and sample are far enough, the unperturbed wave functions of the tip and

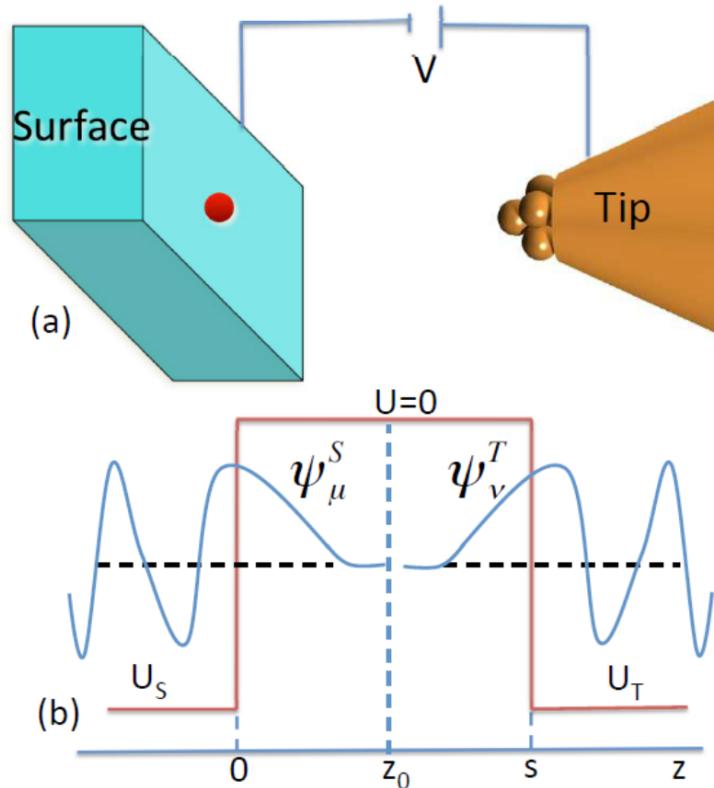


Fig. 1.3 (a) A simplified schematics of the STM system. (b) Corresponding potential and wave function of the tip and sample. This picture is taken from [12]

*⁴ The discussion in this section almost follows a review by Lounis [12]. This review is well summarized and easy to understand. Read it if you are interested.

sample (ψ^T and ψ^S) satisfy the time-dependent Schrödinger equations as follows:

$$i\hbar \frac{\partial \psi^T}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_T \right) \psi^T, \quad (1.10)$$

$$i\hbar \frac{\partial \psi^S}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S \right) \psi^S. \quad (1.11)$$

The solution in the steady state can be expressed by

$$\psi^T = \psi_\nu^T e^{-iE_\nu^T t/\hbar}, \quad (1.12)$$

$$\psi^S = \psi_\mu^S e^{-iE_\mu^S t/\hbar}. \quad (1.13)$$

The time independent Schrödinger equations are:

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_T \right) \psi_\nu^T = E_\nu^T \psi_\nu^T, \quad (1.14)$$

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S \right) \psi_\mu^S = E_\mu^S \psi_\mu^S \quad (1.15)$$

, in which E_ν^T (E_μ^S) are the energy of ν (μ)-th steady state of the tip (sample). After approaching the tip, the potential of the tip and sample merge. Thus, the time evolution of the system is described by a Schrödinger equations using the full potential:

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S + U_T \right) \psi. \quad (1.16)$$

The time evolution of the system can be treated by considering one of the potentials as a perturbation. Assuming that at $t \rightarrow -\infty$, when the tip and sample are far, an electron occupies a steady-state of the sample ψ_μ^S . In addition, we also assume that the tip approaches the sample very slowly compared with the electronic transition. In other words, the approaching process is regarded as adiabatic. This assumption is reasonable because the time scale of electronic transition is fs order, while the approaching is in the order of a second.

At $t \rightarrow -\infty$, the tip and sample are so far that the tip potential does not affect the state of the sample. As the tip approaches, the tip potential gradually affects the sample. This situation can be reproduced by replacing the tip potential in eq. 1.16 with the time-dependent potential as follows:

$$U_T(t) = e^{\eta t/\hbar} U_T, \eta > 0. \quad (1.17)$$

η is a small positive value, and $\eta \rightarrow 0$ at $t \rightarrow \infty$. Thus, $U_T(t) \rightarrow 0$ at $t \rightarrow -\infty$ and $U_T(t) \rightarrow U_T$ at $t \rightarrow \infty$, which represents the potential change during the approaching process. Because of this time-dependent potential, the electron can transfer to the state of the tip ψ^T . This transition is directly related to the tunneling current.

Here we assume that the wavefunction of the whole system (ψ) can be expressed by the linear combination of the unperturbed eigenstates of the tip and sample as follows:

$$\psi = a_\mu(t) \psi_\mu^S e^{-iE_\mu^S t/\hbar} + \sum_{\nu=1}^{\infty} c_\nu(t) \psi_\nu^T e^{-iE_\nu^T t/\hbar}. \quad (1.18)$$

The coefficients ($a_\mu(t)$ and c_ν) are determined by solving the Schrödinger equations with full potential. Because the electron occupies ψ_μ^S at the beginning, $a_\mu(t) = 1$ and $c_\nu(t) = 0$ at $t \rightarrow -\infty$. Here, note that ψ_μ^S and ψ_ν^T are derived from different Hamiltonian. Thus, in general, these states are not orthogonal, and these are not the eigenstate of the combined Hamiltonian. A basic assumption of Bardeen's theory is that the interaction between the tip and sample does not cause any harmonic states. This assumption is valid if the tip-sample distance is not too close. Under this assumption, ψ_μ^S and ψ_ν^T are still the eigenstates even under the combined potential, and these states are orthogonal to each other ($\int \psi_\mu^T \psi_\nu^S d^3r \sim 0$).

Substitution of perturbed wavefunction (eq. 1.18) into the Schrödinger equation (eq. 1.16), and the projection onto ψ_ν^T yields^{*5}

$$i\hbar \frac{dc_\nu}{dt} = \int \psi_\nu^{T*} U_T \psi_\mu^S d^3r e^{-i(E_\mu^S - E_\nu^T + i\eta)t/\hbar} + \sum_{\lambda=1}^{\infty} c_\lambda(t) \int \psi_\nu^{T*} U_S \psi_\lambda^T d^3r e^{-i(E_\lambda^T - E_\nu^T)t/\hbar} \quad (1.19)$$

, or in the bra-ket expression,

$$i\hbar \frac{dc_\nu}{dt} = \langle \psi_\nu^T | U_T | \psi_\mu^S \rangle e^{-i(E_\mu^S - E_\nu^T + i\eta)t/\hbar} + \sum_{\lambda=1}^{\infty} c_\lambda(t) \langle \psi_\nu^T | U_S | \psi_\lambda^T \rangle e^{-i(E_\lambda^T - E_\nu^T)t/\hbar} \quad (1.20)$$

During the calculation, we assumed that $\frac{d}{dt}a_\mu(t) = 0$ and $a_\mu(t) = 1$ because of the adiabatic transition. In addition, $(e^{\eta t/\hbar} - 1) \sim 0$ because this part vanishes at $\eta \rightarrow 0$.

Because c_λ has the value just by the perturbation, it is a small value. Thus, the second term in eq. 1.20 is neglected compared with the first term. Therefore, eq. 1.20 is approximated to be

$$i\hbar \frac{dc_\nu}{dt} = \langle \psi_\nu^T | U_T | \psi_\mu^S \rangle e^{-i(E_\mu^S - E_\nu^T + i\eta)t/\hbar} \quad (1.21)$$

Here the integral $M_{\mu\nu} = \langle \psi_\nu^T | U_T | \psi_\mu^S \rangle$ describes the projection of the initial state ψ_μ^S onto the final state ψ_ν^T , caused by the perturbation (U_T). $M_{\mu\nu}$ is called as matrix element. Integrating by time yields

$$c_\nu = \frac{1}{E_\mu - E_\nu + i\eta} M_{\mu\nu} e^{-i(E_\mu^S - E_\nu^T + i\eta)t/\hbar}. \quad (1.22)$$

The square of c_ν expresses the probability of the transition from the initial state to the final state at t , as follows:

$$|c_\nu(t)|^2 = \frac{e^{2\eta t/\hbar}}{(E_\mu^S - E_\nu^T)^2 + \eta^2} |M_{\mu\nu}|^2 \quad (1.23)$$

The tunneling probability per unit time is expressed as

$$P_{\mu\nu}(t) = \frac{d}{dt} |c_\nu(t)|^2 = \frac{2\eta}{(E_\mu^S - E_\nu^T)^2 + \eta^2} e^{2\eta t/\hbar} \frac{1}{\hbar} |M_{\mu\nu}|^2 \quad (1.24)$$

With using a definition of the delta function ($\delta(x) = \frac{1}{\pi} \lim_{\eta \rightarrow 0} \frac{\eta}{x^2 + \eta^2}$), eq. 1.24 at $\eta \rightarrow 0$ is expressed as follows:

$$P_{\mu\nu}(t) = \frac{2\pi}{\hbar} \delta(E_\mu^S - E_\nu^T) |M_{\mu\nu}|^2 \quad (1.25)$$

This is the general result of the first-order perturbation theory, called Fermi's golden rule. The tunneling current is proportional to $P_{\mu\nu}(t)$ multiplied by electron charge (e).

1.4.2 Tunneling current

In the previous section, we only consider the transition from a state of the sample ψ_μ^S to those of the tip ψ_ν^T . However, because the energy state of the tip and sample consists of a spectrum, all the possible transitions should be considered to get the net tunneling current. The electron transfer happens from the filled state to the empty state. The Fermi-Dirac distribution describes the energy distribution of electrons as follows:

$$f(E - E_F) = \frac{1}{1 + \exp[(E - E_F)/k_B T]} \quad (1.26)$$

, where E_F represents the Fermi level. For example, a state ψ_μ^S at the energy of E_μ^S is occupied with the probability of $f(E_\mu^S - E_F)$, or unoccupied with the probability of $1 - f(E_\mu^S - E_F)$. If we assume that the

^{*5} See appendix for detailed calculation process.

bias V is applied between the tip and sample, the tunneling current from the sample to tip ($I_{S \rightarrow T}$) and the tip to sample ($I_{T \rightarrow S}$) is expressed as follows:

$$I_{S \rightarrow T} = \frac{4\pi e}{\hbar} \sum_{\mu\nu} f(E_\mu^S - E_F^S)[1 - f(E_\nu^T - E_F^T)]|M_{\mu\nu}|^2 \delta(E_\mu^S - E_\nu^T - eV), \quad (1.27)$$

$$I_{T \rightarrow S} = \frac{4\pi e}{\hbar} \sum_{\mu\nu} f(E_\nu^T - E_F^T)[1 - f(E_\mu^S - E_F^S)]|M_{\mu\nu}|^2 \delta(E_\mu^S - E_\nu^T - eV). \quad (1.28)$$

Note that a factor of 2 is multiplied considering the spin degree of freedom. Thus, the net tunneling current is the difference between these two:

$$I = I_{S \rightarrow T} - I_{T \rightarrow S} = \frac{4\pi e}{\hbar} \sum_{\mu\nu} [f(E_\mu^S - E_F^S) - f(E_\nu^T - E_F^T)]|M_{\mu\nu}|^2 \delta(E_\mu^S - E_\nu^T - eV). \quad (1.29)$$

In eq. 1.29, all the combination of μ and ν are considered. However, because of the δ function, just considering the transition with an energy difference $E_\mu^S = E_\nu^T + eV$ is enough. Therefore, summarizing the states as the density at the energy level, eq. 1.29 can be written as follows:

$$I = \frac{4\pi e}{\hbar} \int d\epsilon [f(E_F^T - eV + \epsilon) - f(E_F^S + \epsilon)] \times n^T(E_F^T - eV + \epsilon) n^S(E_F^S + \epsilon) |M(E_F^S + \epsilon, E_F^T - eV + \epsilon)|^2 \quad (1.30)$$

$n^T(E)$ and $n^S(E)$ are called the density of states (DOS). If $T = 0$ or the thermal energy ($k_B T$) is much smaller than the experimental resolution, eq. 1.30 is simplified as

$$I = \frac{4\pi e}{\hbar} \int_0^{eV} d\epsilon n^T(E_F^T - eV + \epsilon) n^S(E_F^S + \epsilon) |M|^2. \quad (1.31)$$

These equations indicate that the tunneling current depends on the DOS of the tip and sample and the transfer matrix.

1.4.3 Bardeen's tunneling matrix

We saw that the tunneling current depends on the tunneling matrix, which definition is given by

$$M_{\mu\nu} = \int_{z>z_0} \psi_\nu^{T*} U_T \psi_\mu^S d^3r. \quad (1.32)$$

Because $U_T = 0$ at the sample side, the integration is performed only at the tip side. We will further consider the expression of the tunneling matrix. With using unperturbed Schrödinger equation (eq. 1.14 and 1.15) and considering the elastic tunneling ($E_\mu^S = E_\nu^T$) and $U_S = 0$ in the integrated region, the matrix is converted as follows:

$$\begin{aligned} M_{\mu\nu} &= \int_{z>z_0} (\psi_\nu^{T*} U_T) \psi_\mu^S d^3r \\ &= \int_{z>z_0} \psi_\mu^S (E_\nu^T + \frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2}) \psi_\nu^{T*} d^3r \\ &= \int_{z>z_0} (\psi_\mu^S E_\mu^S \psi_\nu^{T*} + \frac{\hbar^2}{2m} \psi_\mu^S \frac{\partial^2 \psi_\nu^{T*}}{\partial z^2}) d^3r \\ &= -\frac{\hbar^2}{2m} \int_{z>z_0} (\psi_\nu^{T*} \frac{\partial^2 \psi_\mu^S}{\partial z^2} - \psi_\mu^S \frac{\partial^2 \psi_\nu^{T*}}{\partial z^2}) d^3r \end{aligned} \quad (1.33)$$

With using the relation

$$\psi_\nu^{T*} \frac{\partial^2 \psi_\mu^S}{\partial z^2} - \psi_\mu^S \frac{\partial^2 \psi_\nu^{T*}}{\partial z^2} = \frac{\partial}{\partial z} [\psi_\nu^{T*} \frac{\partial \psi_\mu^S}{\partial z} - \psi_\mu^S \frac{\partial \psi_\nu^{T*}}{\partial z}] \quad (1.34)$$

, integration over the z direction can be carried out to obtain

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int_{z=z_0} [\psi_\nu^{T*} \frac{\partial^2 \psi_\mu^S}{\partial z^2} - \psi_\mu^S \frac{\partial^2 \psi_\nu^{T*}}{\partial z^2}] dx dy. \quad (1.35)$$

This is the one-dimensional form of Bardeen's tunneling matrix, expressed by a surface integral of the wave functions at the interface. This expression clarifies that the tunneling current is also affected by the spatial distribution of the wave functions of the tip and sample.

1.4.4 Tersoff-Hammann model

According to Bardeen's theory, we need information about the electronic structure of the tip and the sample for evaluating the tunneling current. It is possible by using sophisticated DFT calculations if the atomic structure of the tip is known, but the determination of the tip shape is not a trivial task. Therefore, a simple model to express the electronic structure of the tip was proposed, which is called as Tersoff-Hammann (TH) model.

(Not yet)

Chapter 2

Lock-in amplifier

A direct way to get the dI/dV spectra is numerical differentiation of the I-V curve. However, the differential process enhances the noise in the I-V curve. Another smoothing process to remove the noise is necessary, leading to broadening the spectral feature. An alternative method to obtain the dI/dV signal is to use a lock-in amplifier. The lock-in amplifier enhances the signal with a specific frequency and phase, effectively canceling the noises. As a result, low noise measurement is possible. In addition, because it is not necessary to sweep the bias in lock-in detection, simultaneous measurement of dI/dV and topography is realized, which visualizes the spatial distribution of the LDOS at the fixed bias. In this chapter, the basic principles of lock-in detection will be introduced [4–7].

2.1 Basic principle

A modulation bias is added to the bias ($V \rightarrow V + \Delta V \sin(\omega t)$) to obtain the dI/dV signal in lock-in detection^{*1}. The additional modulation bias causes the oscillation in the tunneling current, which can be expanded by the Taylor series as follows:

$$I(V + \Delta V) = I(V) + \frac{dI}{dV}(\Delta V \sin \omega t) + \frac{d^2 I}{dV^2} \frac{1}{2}(\Delta V \sin \omega t)^2 + \dots \quad (2.1)$$

Then, the current signal is multiplied by a reference wave $A \sin(\omega t + \theta)$. The first and second terms of the product are

$$\text{First term} \quad I(V) \times A \sin(\omega t + \theta) \quad (2.2)$$

$$\begin{aligned} \text{Second term} \quad & \frac{dI}{dV}(\Delta V \sin \omega t) \times A \sin(\omega t + \theta) \\ &= \frac{dI}{dV} \frac{A \Delta V}{2} \{ \cos \theta + \sin(2\omega t + \theta) \} \end{aligned} \quad (2.3)$$

The components of these terms can be classified into two kinds: constant and oscillating terms. The oscillating components are removed by applying a low-pass filter (LPF). As a result, the remaining signal is

$$V_X = \frac{dI}{dV} \frac{A \Delta V}{2} \cos(\theta). \quad (2.4)$$

Finally, the remaining signal (V_X) is proportional to the dI/dV . V_X is called an in-phase signal because the value is maximized when the modulation and reference waves have the same phase ($\theta = 0$). If $A \cos(\omega t + \theta)$ is used as a reference, the output after passing the LPF is

$$V_Y = \frac{dI}{dV} \frac{A \Delta V}{2} \sin(\theta), \quad (2.5)$$

, which is maximized when the phase is 90° rotated from the reference. Because the output signal depends on the phase of the reference wave, the correct setting of the phase is crucial in lock-in measurement. On the other hand, a phase insensitive value can be obtained from V_X and V_Y as follows:

$$V_R = \sqrt{V_X^2 + V_Y^2} = \frac{A \Delta V}{2} \left| \frac{dI}{dV} \right|. \quad (2.6)$$

^{*1} See appendix for the method to get the second derivative

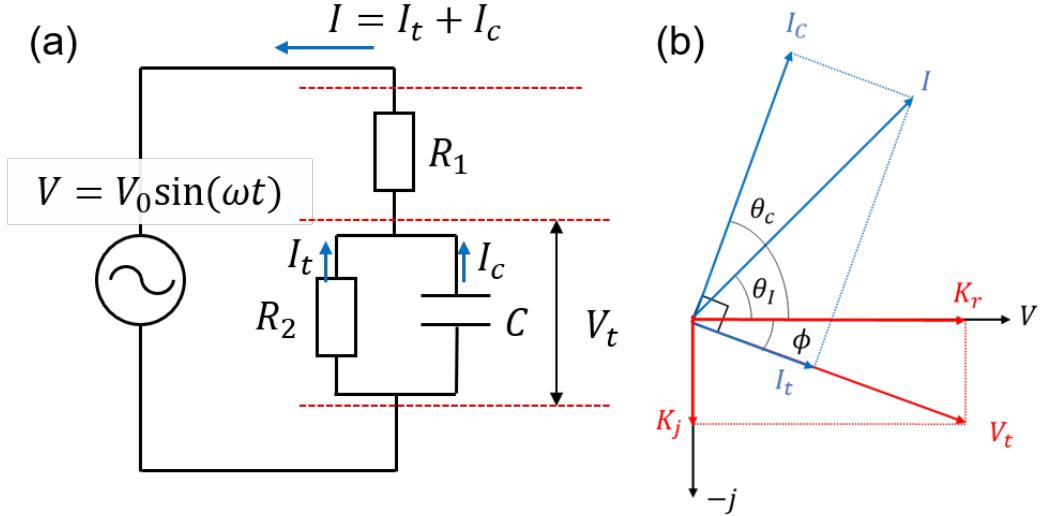


Fig. 2.1 (a) Electric circuit corresponding to STM. R_1 is the resistance of cables, and R_2 and C is the resistance and capacitance of the tunneling part. (b) Relationship of phases of signals.

The dI/dV signal is obtained from V_R without caring about the phase of the reference wave. However, this signal loses the phase sensitivity, possibly resulting in a larger noise.

2.2 Cancellation of noise by lock-in detection

Let us consider how the noise signal in the tunneling current is treated in the lock-in process. Assuming that the noise signal is expressed as $B \sin(\omega_1 t + \phi)$, the product with the reference wave is as follows:

$$B \sin(\omega_1 t + \phi) \times A \sin(\omega t + \theta) = AB[\cos\{(\omega_1 - \omega)t + \theta - \phi\} - \cos\{(\omega_1 + \omega)t - (\theta - \phi)\}] \quad (2.7)$$

If $\omega_1 - \omega$ is larger than the cutoff frequency of LPF, the noise is effectively eliminated in the output signal. Besides, the noise with the phase other than $\phi = \theta$ is attenuated. As a result, almost all the noise except for $\omega_1 = \omega$ and $\phi = \theta$ is canceled out by the lock-in processes. Therefore, a simple way to reduce the noise is to choose ω with the smallest noise. It is recommended to check the noise in tunneling current before setting ω and avoid the frequency with the large noise. Especially, ω should not be set at the multiples of 60 Hz because these frequencies often contain large noise originating from the electrics.

2.3 Setting the phase in STS measurement

The modulated tunneling current has a different phase from the modulation bias because of the phase shift in the circuit. Therefore, the largest dI/dV signal can not be obtained by simply setting $\theta = 0$ (same phase with the modulation). In this section, the relationship of the phase of the signals will be discussed. Then a method to set a proper phase will be introduced.

Figure 2.1 shows the simplified electric circuit corresponding to the STM system. The impedance of the tunneling part is expressed as follows by using the imaginary unit (j)^{*2}:

$$Z_t = \frac{1}{\frac{1}{R_2} + j\omega C} \quad (2.8)$$

The total impedance of the circuit is

$$Z = R_1 + Z_t. \quad (2.9)$$

^{*2} Impedance is the value corresponding to resistivity. In the alternating circuit, the capacitor can be treated as a resistor with the resistance of $\frac{1}{j\omega C}$. Note that the resistor (R_2 and capacitor are parallel, and the tunneling part and resistor (R_2) are series connections.

The bias applied to the tunneling part is calculated as follows:

$$\begin{aligned} V_t &= V \times \frac{Z_t}{Z} \\ &= \frac{VR_2(R_1 + R_2)}{(R_1 + R_2)^2 + (R_1 R_2 \omega C)^2} - j \frac{VR_1 R_2^2 \omega C}{(R_1 + R_2)^2 + (R_1 R_2 \omega C)^2} \\ &= K_r - jK_j \end{aligned} \quad (2.10)$$

While K_r has the same phase with respect to the alternative current (AC) source, the imaginary component (K_j) is 90° delayed. The relationship of the phases of the signals is summarized in Fig. 2.1(b). One can see that the other expression of V_t is

$$V_t = \sqrt{K_r^2 + K_j^2} \sin(\omega t - \phi). \quad (2.11)$$

$$\phi = \tan^{-1} \frac{K_j}{K_r} = \tan^{-1} \frac{R_1 R_2 \omega C}{R_1 + R_2}. \quad (2.12)$$

Equation 2.11 indicates that the phase of bias applied on the tunneling part delays by ϕ with respect to the modulation bias.

Then, we should consider the phase of the tunneling current with respect to V_t . The current in the circuit (I) consists of two components; tunneling current (I_t) and capacitive current(I_C). The relationships between the applied bias (V_T) and currents (I_t and I_C) are expressed as follows:

$$I_t = \frac{V_t}{R_2} \quad (2.13)$$

$$I_c = \frac{dQ}{dt} = C \frac{dV_t}{dt} \quad (2.14)$$

While I_t has the same phase as V_t , the phase of I_c is 90° in advance. These relationships are also summarized in Fig. 2.1(b). Finally, I_t , observable in STS measurement, is delayed by ϕ from the modulation bias. Because the lock-in measurement is phase-sensitive, we should set the phase of reference wave $\theta = \phi$ for getting the largest signal. The following process can achieve this phase setting *³:

1. Retract the tip until the tunneling current becomes zero. Even after retracting, I_c still survives because the capacitance of the tunneling part causes it. Therefore, $I = I_c$.
2. Set the phase to get the largest signal (Press “Auto phase”). Now the phase is set to $\theta = \theta_C$.
3. Rotate the phase by -90° . Because the tunneling current is delayed by 90° from I_c , now the phase is set to $\theta = \phi$.

The phase insensitive output (V_R) can also be used in STS. In this case, the output reflects the value of I without depending on the phase. Therefore be careful that the I_c component act as an offset in the measurement *⁴.

2.4 LPF and time constant

A key component of lock-in detection is the LPF. Although the measurement with lower noise is possible by using LPF with lower cut-off frequency (and higher roll-off, as described below), the response of the system gets slower instead. In this section, the basic principle and properties of LPF will be reviewed.

The most simple low-pass circuit is shown in Fig. 2.2. The relationship between the input signal V_i with the frequency of ω and the output signal V_O is expressed as follows:

$$G = \frac{V_0}{V_i} = \frac{1}{\sqrt{1 + (\omega CR)^2}}. \quad (2.15)$$

*³ Sometimes the STM controller have the additional phase shift of π . Make sure that the phase setting is proper in your STM by measuring dI/dV.

*⁴ I always use V_R without considering the phase. Although it was enough for my experiment, proper phase setting can make the data higher quality.

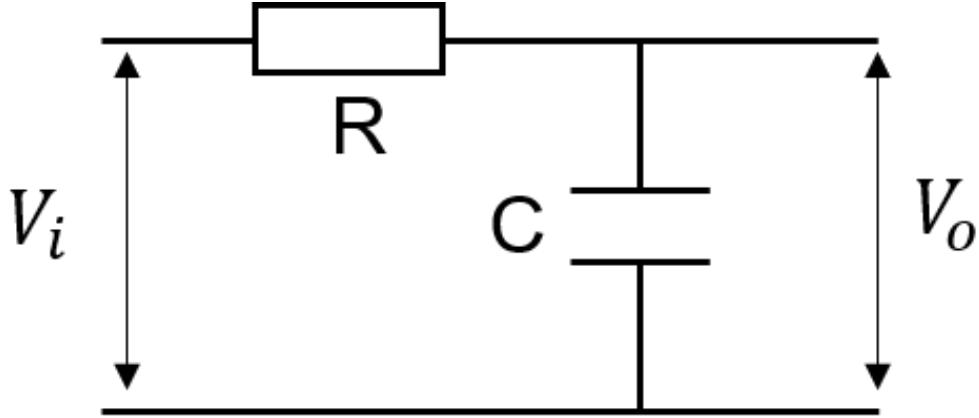


Fig. 2.2 Low pass filter

The red curve in Fig. 2.3(a) shows the gain G as a function of frequency ^{*5}. This circuit effectively attenuates the signal with the frequency above $\omega \sim 1$, which is the main characteristic of LPF. The cutoff frequency of the LPF is defined as the frequency where the gain = -3 dB (corresponding to $V_o = V_i/\sqrt{2}$). Therefore, $\omega_c = 1/RC$ or $f = 1/(2\pi RC)$ (Hz), indicating that the higher RC realizes the lower cutoff.

The LPF can be multiplied to enhance the cutoff ability, as shown in the inset of Fig. 2.3(a). This circuit is called a cascade. The gain of n_{th} cascade is expressed as follows:

$$G_n = \left\{ \frac{1}{\sqrt{1 + (\omega CR)^2}} \right\}^n \quad (2.16)$$

The total gain of $n = 1 \sim 4$ th cascade is shown in Fig. 2.3. The cascade with larger n shows the steeper decay above ω_c . The cutoff ability of LPF is evaluated by the gain change when $\omega_c \rightarrow 2\omega_c$. This value is called roll-off. The roll-off of n_{th} cascade is expressed as $6n$ (dB/oct.). ^{*6}

Although the performance of LPF becomes better by using higher RC and more cascades, it makes the response time slower. The following equation expresses the response time of the n_{th} LPF:

$$\begin{aligned} V &= V_{max} \left\{ 1 - e^{-t/RC} \sum_{k=0}^{n-1} \frac{1}{k!} \left(\frac{t}{RC} \right)^k \right\} \\ &= V_{max} \left\{ 1 - e^{-t/\tau} \sum_{k=0}^{n-1} \frac{1}{k!} \left(\frac{t}{\tau} \right)^k \right\} \end{aligned} \quad (2.17)$$

Figure 2.3(b) shows the change of the output according to eq. 2.17. The time axis is normalized by a value called the time constant ($\tau = RC$). The figure indicates that the output saturates after several times the time constant. In addition, the cascade takes a much longer time to get the final value. The relationship of cascade number, roll-off, and the time to get the output is summarized in table 2.1 [6]. In general, the LPF with better performance (low cut-off and high roll-off) takes a longer time to reach the final value.

A map of DOS on the surface is obtained by recording the lock-in output during the scanning. This measurement is called dI/dV mapping or STS mapping. In the mapping measurement, the output from lock-in is recorded in each pixel of the image. If the recording time for a pixel is too short, the lock-in output can not follow the change of the observable (dI/dV). For example, consider that the scanning is performed with the scanning speed of 2 s/line with 256 pixels. The measurement time for a pixel is $2/256 \sim 7$ ms. If the time constant and roll-off are 30 ms and 6 dB/oct, the output reaches only 21 % of the maximum value. For getting clearer data, a much longer scanning time is recommended.

^{*5} The unit "dB" is defined as $20 \log \frac{V_o}{V_i}$

^{*6} oct: short expression of octave, which means the frequency is doubled

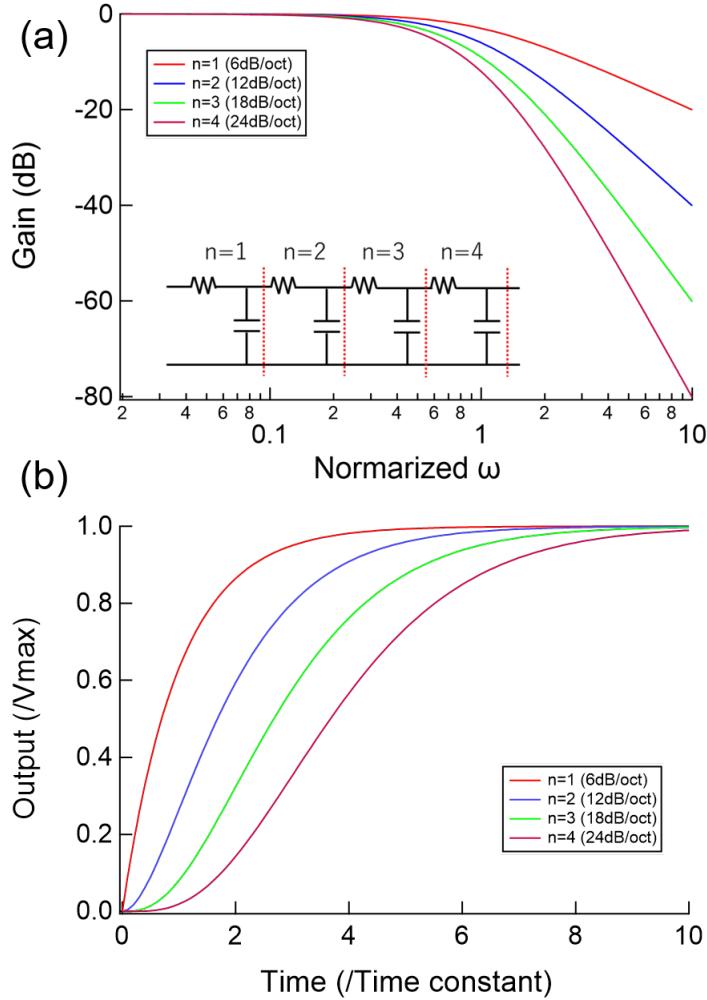


Fig. 2.3 (a) Gain as a function of normalized frequency (CR=1) (b) time dependence of output

Table 2.1 Time to reach 63.2, 90 and 99 % of final values for n_{th} cascade.

Order n	Roll-off dB/oct	setting times (/TC)		
		63.2%	90%	99%
1	6	1	2.3	4.6
2	12	2.2	3.9	6.6
3	18	3.3	5.3	8.4
4	24	4.4	6.7	10

2.5 Broadening of spectral feature

While the S/N becomes better when using a larger modulation, the spectral feature becomes broader, resulting in poor energy resolution. It can be intuitively understood by considering that the larger modulation indicates electrons with a broader energy range affects the measurement. By combining the thermally induced broadening, the broadening of the spectral feature is given by the following equation [8]:

$$\Delta E = \sqrt{(3.3k_B T)^2 + (2.5V_{mod})^2}. \quad (2.18)$$

An example of the broadening of the spectral feature induced by the modulation is shown in Fig. 2.4 [9]. Figure 2.4 shows the STS spectra of the superconducting gap of Pb(111) with various modulation biases.

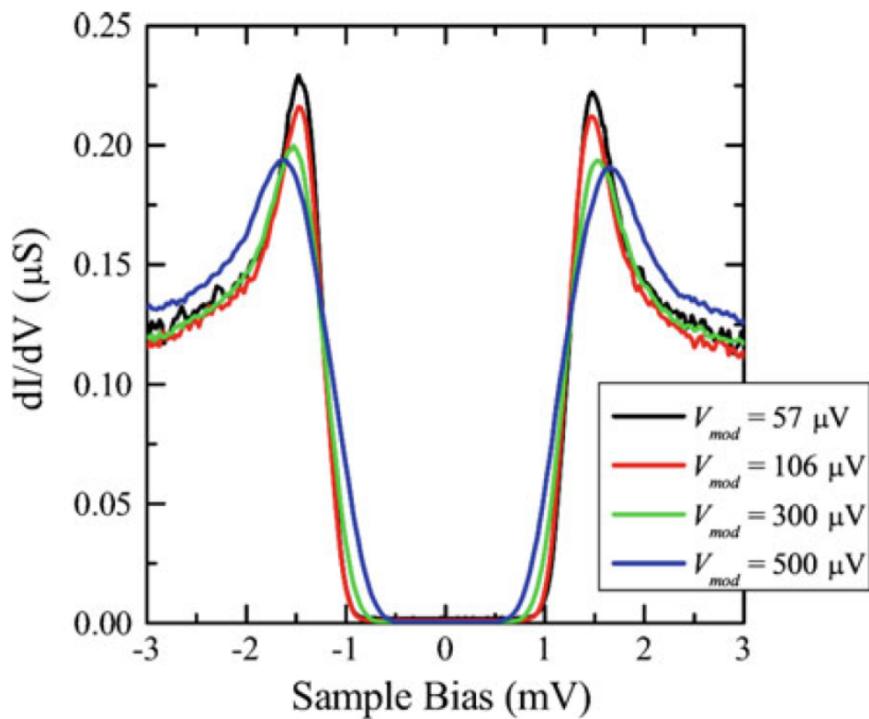


Fig. 2.4 Superconducting gap of Pb(111) measured at 0.4 K by STS with various modulation bias [9].

The slope near the edge of the superconducting gap becomes steeper as the smaller modulation bias.

Chapter 3

Distortion of STM images

3.1 Thermal drift

One of the conventional characteristics of solids is thermal expansion—the volume of the solids changes when temperature varies. The volume change is so tiny that no one cares about it in daily life. However, thermal expansion is visible in STM observation. Because the components in the STM system differently expand, the relative position of the tip and sample shifts by the temperature change, as shown in Fig. 3.1(a). This effect is called thermal drift. Figure 3.1(b) illustrates the effect of thermal drift on scanning. Although the scanning controller moves the tip to scan a square region, the scanning region is a parallelogram because of the additional position change by thermal drift. Because the parallelogram scanning region is recorded as the square image, the features in the image distort, as shown in Fig. 3.1(c). The image distortion leads to misunderstanding the properties of the sample, such as lattice constant and symmetry of the lattice. In addition, spectroscopic observation with a fixed tip position is difficult if thermal drift is present. Although there are several methods to avoid thermal drift, the easiest way is to wait until the temperature becomes stable.

The best way to avoid the thermal drift is waiting. However, sometimes even a minimal drift becomes apparent, especially when long time scanning is necessary. In that case, it would be better to consider removing the distortion by numerical process on the image. Here a method to calculate the speed of drift and recover the undistorted image will be introduced.

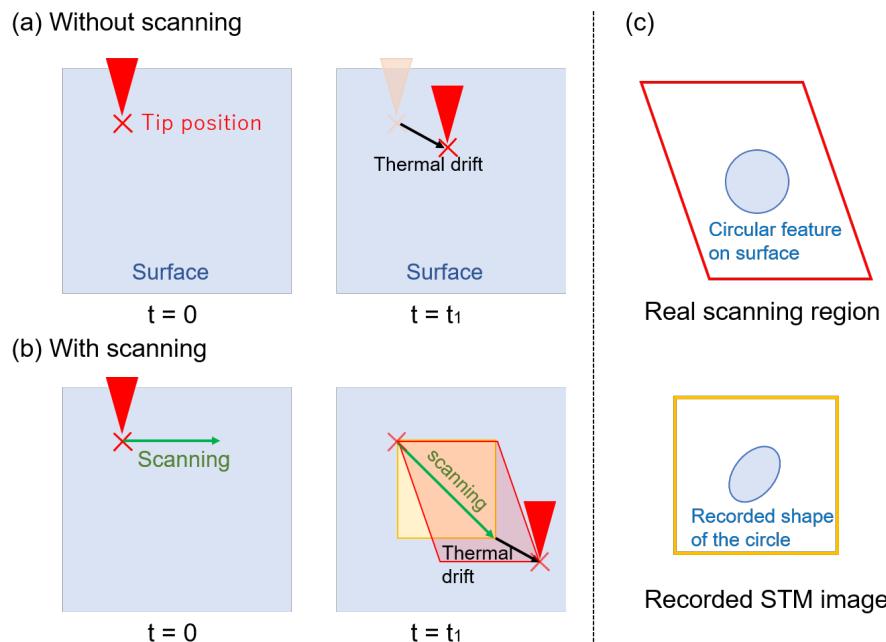


Fig. 3.1 (a) The change of relative position of the tip and sample. (b) The relationship between the STM scan and thermal drift. (c) The effect of thermal drift on STM image.

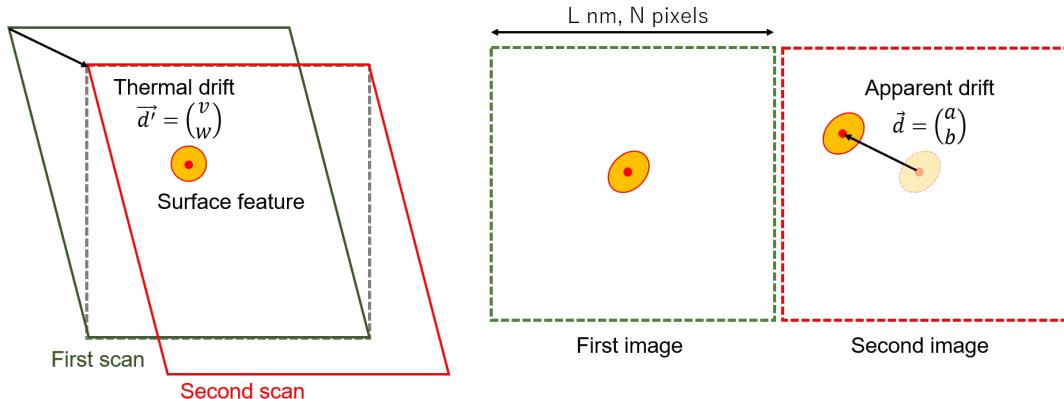


Fig. 3.2 Schematic illustration of the effect of thermal drift. The value of thermal drift can be estimated from the two images which are taken subsequently in the same region.

The speed of thermal drift can be calculated from successively obtained two images (the second image should be taken just after the first image without changing any parameters.). The position of the surface feature (such as defects) shifts because of thermal drift, as shown in Fig. 3.2. This shift is called apparent drift. The actual value of the thermal drift is calculated from the apparent drift as follows^{*1}:

$$v = -\frac{a}{1 + a/2NL + b/L} \quad (3.1)$$

$$w = -\frac{b}{1 + a/2NL + b/L} \quad (3.2)$$

With using this actual drift value, the transformation matrix to recover the undistorted image is expressed as follows:

$$A = \frac{1}{L} \begin{pmatrix} L + v/2N & v \\ w/2N & L + w \end{pmatrix}$$

By applying this matrix, a position \vec{r} in the original image is converted to the position in the undistorted image by $\vec{r}' = A\vec{r}$.

The other method uses the symmetry in the STM image. [To be written]

3.2 Creep

The other major source of the distortion is the creep effect of the piezoelectric^{*2}. This is a relatively short time effect compared with the thermal drift. Assuming that a step-like bias voltage is applied to the piezo, as indicated by the blue curve in Fig. 3.3(a). Although the applied bias immediately reaches the set value, the deformation of the piezoelectric (or tip position) does not follow simultaneously. Instead, the shape of the piezo gradually changes as the red curve in Fig. 3.3(a). This is called the creep effect. The creep distorts the image, as in Fig. 3.3(b). Assume that the tip is moved to X-direction (a bias is applied to the piezo to move the tip) and start scanning just after the movement. Because the position of the tip cannot follow the change of the bias immediately, the scanning starts from the displaced position. As the scanning proceeds, the scanning line gradually approaches the intended scanning region. As a result, the actual scanning region is out of the intended region, which distorts the image. The distortion caused by creep becomes apparent significantly after the position of the tip changed drastically.

The creeping also affects the vertical direction (Z). Especially, careful care is necessary when the feedback is off, such as STS measurement. Figure 3.4 shows the motion of the tip when offset bias along the Z direction is applied to the piezo. The tip moves to the sample surface with creeping. If the tip

^{*1} See appendix if you are interested in.

^{*2} Piezoelectric: the shape of the material changes with applied voltage. Piezoelectric material is used to move the tip (or sample) in STM.

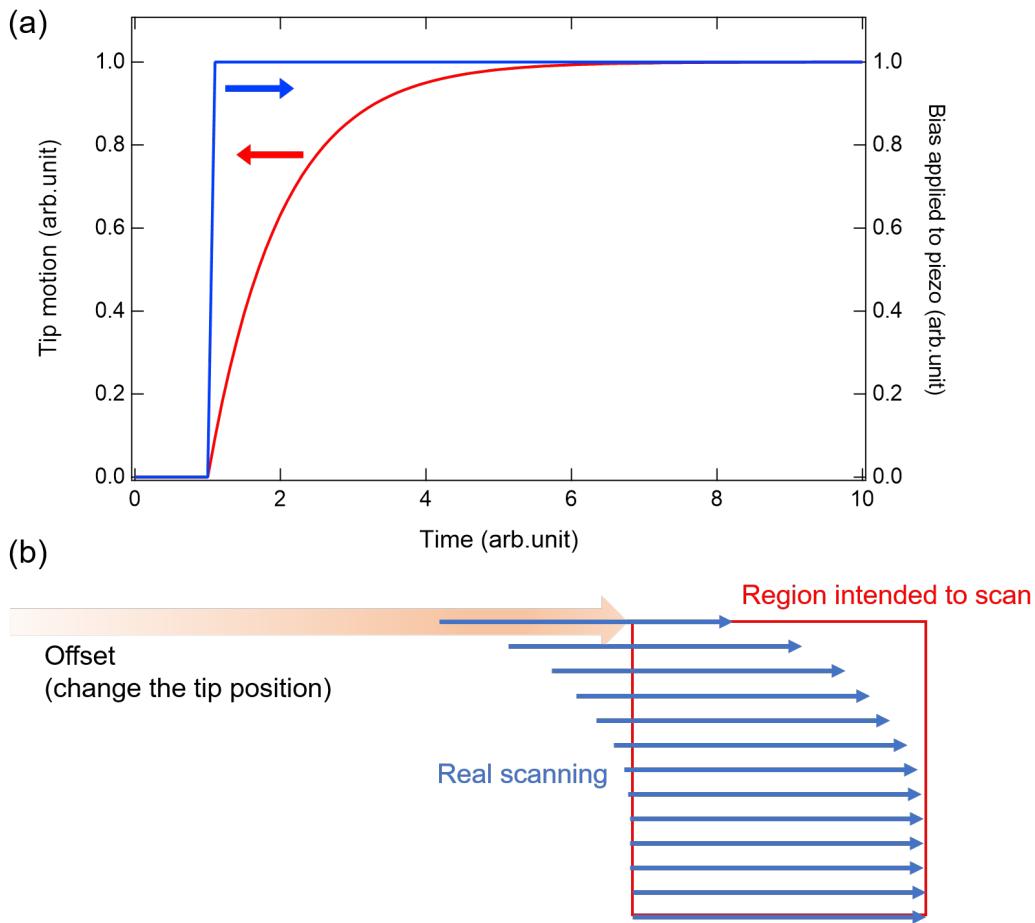


Fig. 3.3 (a) The response of the piezo to the applied bias voltage. (b) The effect of creeping on the scanning region.

detects the tunneling current, the feedback loop tries to keep the tip height constant. Thus, the motion of the tip by creeping is canceled by the feedback motion. Here, assuming that the feedback loop is switched off for STS measurement. Because the feedback motion does not cancel the creeping effect anymore, the tip starts to move toward the sample, resulting in the crash. Therefore, before turning off the feedback, wait for enough time to suppress the creeping effect.

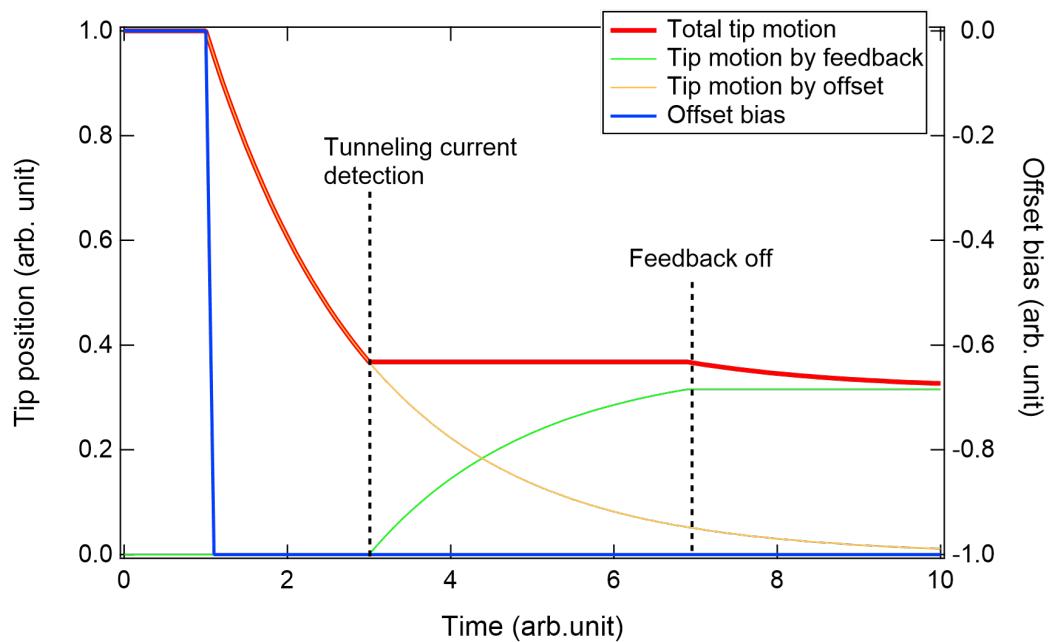


Fig. 3.4 The effect of creeping on the vertical position of the tip.

Chapter 4

Interference of electrons on surfaces

During the STM observation, one may have a chance to observe a wavy pattern around the defects, steps, and adsorbates. This wavy pattern is caused by the interference of incident and scattered electron waves, called quasi-particle interference (QPI). Because the scattering of electrons affects the resistance, thermal electricity, magnetic susceptibility, and electron transition, the detailed consideration of the scattering event should help understand the characteristics of the materials. In addition, the wavenumber of the standing wave correlates with the band structure of the crystals. Thus, the information of the band structure can be deduced from the QPI, like ARPES experiments.

In this chapter, the basic mechanism of QPI is discussed based on a simple one-dimensional case. In addition, the relationship between the band structure and the QPI pattern is introduced.

4.1 Interference of one-dimensional case

The wave function of an electron with the wavenumber of k , the energy of E , and a phase factor θ is expressed as

$$\phi = A \exp(ikx) \times \exp(-\frac{iEt}{\hbar}) \times \exp(i\theta). \quad (4.1)$$

Let us assume the 1D scattering as shown in Fig. 4.1. An incident wave with the wavenumber of k_1 , energy of E_1 , and phase of θ_1 is scattered by a potential barrier. After the scattering, the wave shows the wavenumber of k_2 and energy of E_2 , and the scattering causes the phase shift of $\delta\theta$. Thus, the whole wave function of the system is expressed as follows:

$$\Phi = B \exp(ik_1 x) \exp(-\frac{iE_1 t}{\hbar}) \exp(i\theta_1) + C \exp(ik_2 x) \exp(-\frac{iE_2 t}{\hbar}) \exp(i(\theta_1 + \delta\theta)) \quad (4.2)$$

, where B and C represent the amplitude of the waves. The observable in the experiment is the square

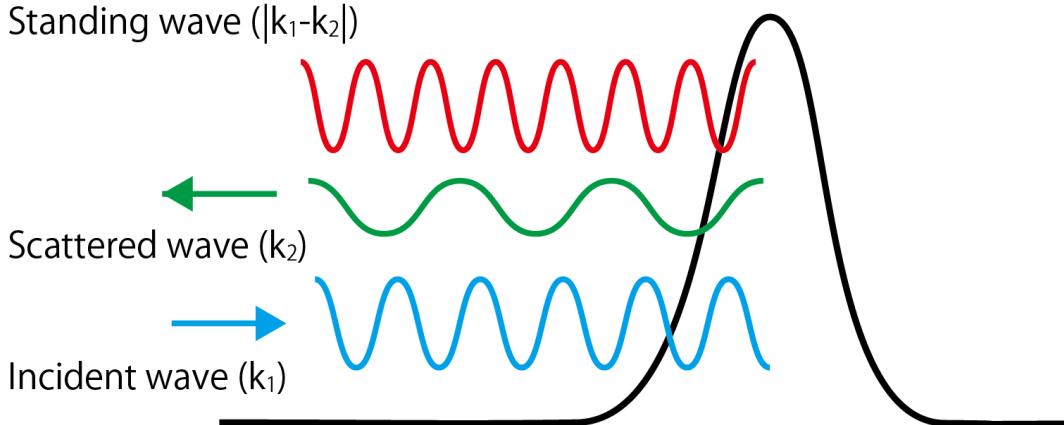


Fig. 4.1 Schematic image of one-dimensional scattering.

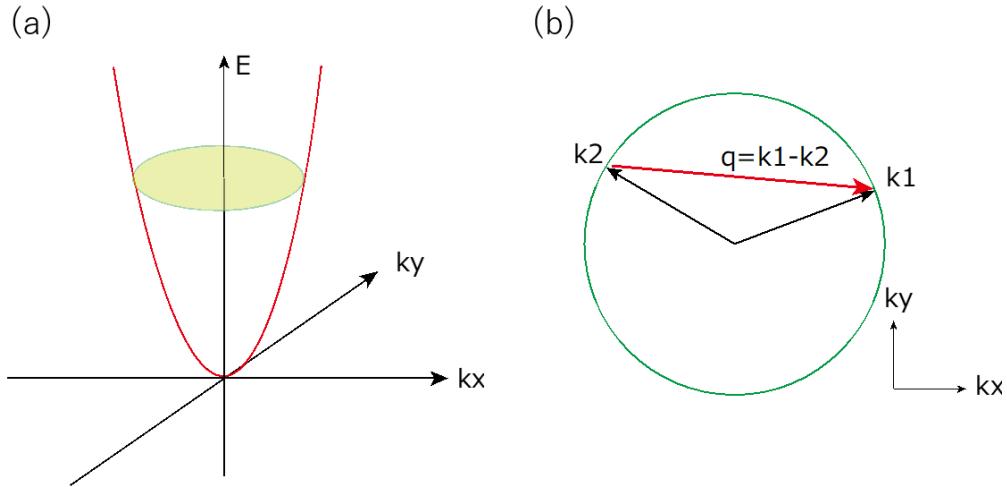


Fig. 4.2 (a) Band structure of two-dimensional free electron. (b) Constant energy contour of the band shown in (a).

of the wave function as follows:

$$|\Phi|^2 = B^2 + C^2 + 2BC\cos[(k_1 - k_2)x - \frac{(E_1 - E_2)}{\hbar}t + \delta\theta] \quad (4.3)$$

The time-dependent term cannot be observed in STM because of the averaging of the signal over the scanning time. Thus, the observable in STM is the time-independent term, which appears when considering the elastic scattering ($E_1 = E_2$). The equation can be written as

$$|\Phi|^2 = B^2 + C^2 + 2BC\cos[(k_1 - k_2)x + \delta\theta]. \quad (4.4)$$

This time-independent wave is called a standing wave, and the formation of the standing wave is caused by the QPI. This standing wave is observable in STM.

This simple model reveals several essential characteristics of the standing wave in STM observation. First, the standing wave does not depend on the phase of the incident wave. Although many electrons with various phases move around the surface, all the incident electrons have the counterpart with the phase shift of $\delta\theta$. Second, only elastic scattering causes the standing wave. In crystals, the possible wavenumber of the electrons is determined by the band structure. Because of the elastic scattering condition, possible k_1 and k_2 should be found on the contour of the band structure at a specific energy.

4.2 QPI and band structure

The wavenumber of the standing wave is expressed as $q = k_1 - k_2$. In the crystals, the possible wavenumber of electrons is determined by the band structure. Here we assume a simple free-electron like band, which has the parabolic shape expressed by

$$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) \quad (4.5)$$

, where m^* is the effective mass. The schematic of the band structure is shown in Fig. 4.2(a). The constant energy cut (CEC) of this band at certain energy forms a circle, as shown in Fig. 4.2(b). The wavenumber of electrons before and after the scattering (k_1 and k_2) should be located on this circle. The wavenumber of a standing wave is determined by the wavenumber of the incident and scattered wave as $q = k_1 - k_2$, which is schematically indicated by the red arrow in Fig. 4.2(b). Considering that the incident and scattered electrons can have any states on the circle, q with all the combinations of k_1 and k_2 are possible. Thus, in the experiment, all the possible q are observed simultaneously. Fig. 4.3 shows the distribution of q from the circular constant energy cut (this is called QPI pattern). The q plot shows a circular distribution with a radius twice larger than the original circle ($|q| = 2|k|$).

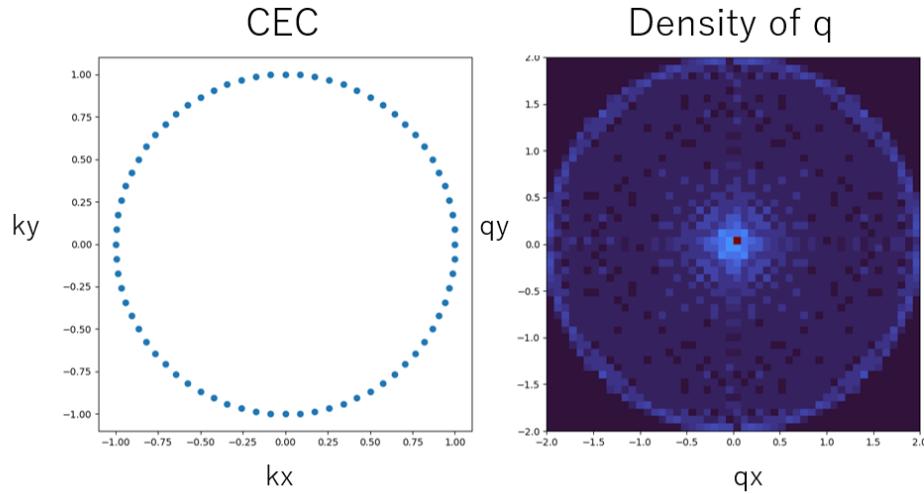
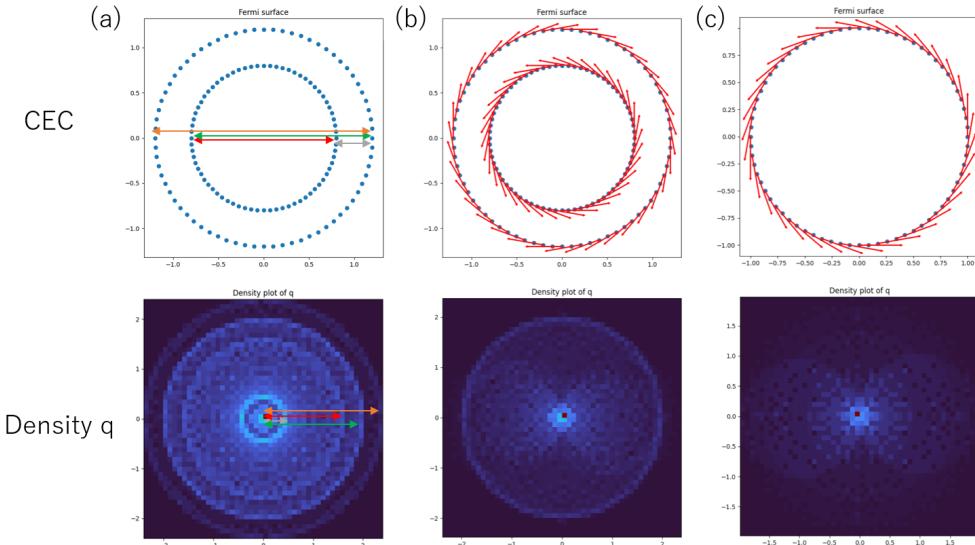
Fig. 4.3 Density of q estimated from the circular CEC.

Fig. 4.4 Predicted QPI pattern from various CEC. (a) Two circular CEC (b) Two circular CEC with spin (c) One circular CEC with spin.

The above single circular CEC case is the most simple one. Here let us try to consider complicated cases, the multiple bands, and with spins. Figure 4.4(a) shows the CEC with two circles. There are four kinds of possible back scatterings, indicated by the arrows. As a result, the q pattern consists of four circles, while the original CEC has only two circles. A general misunderstanding of the QPI pattern is that the QPI pattern directly represents the shape of CEC. However, this example shows that, in general, the CEC and QPI are not identical. Therefore, careful consideration is necessary to analyze the relationship between the CEC and QPI patterns.

The CEC in Fig. 4.4(b) also consists of two circles, but the circles are spin-polarized with opposite chirality as the spin direction indicated by the red arrows ^{*1}. Because the states with opposite spin are orthogonal to each other, these states do not interfere (in other words, the scattering is forbidden). As

^{*1} This situation is realized by Rashba splitting

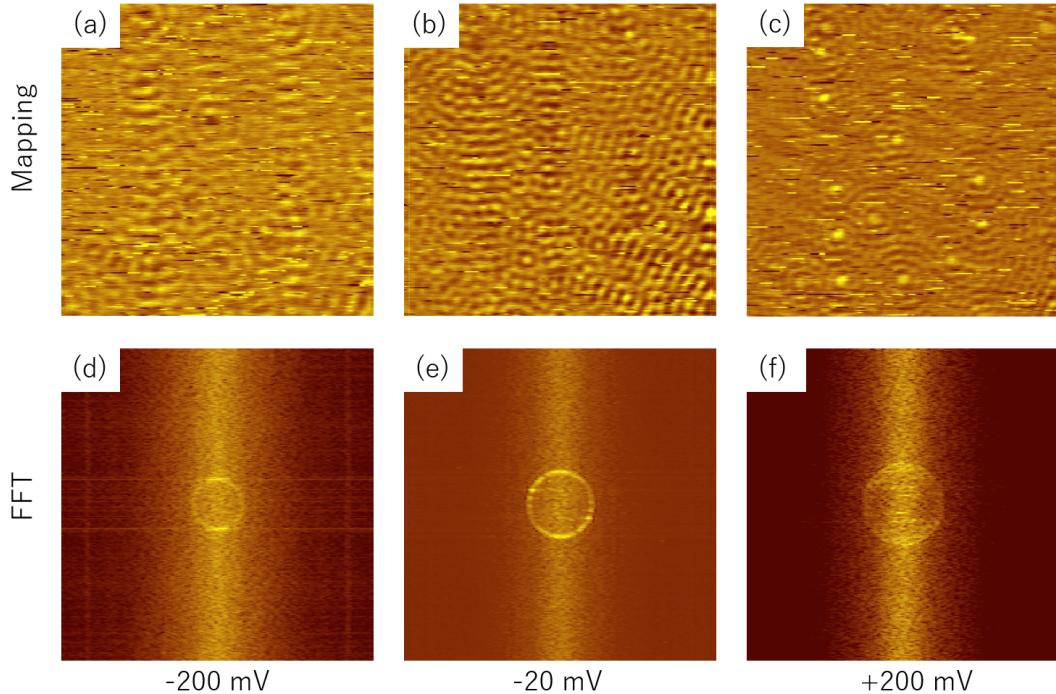


Fig. 4.5 (a)-(c) dI/dV image ($50 \times 50 \text{ nm}^2$, 1 nA) taken at -200, -20, and +200 mV, respectively. Corresponding FFT images are shown in (d)-(f).

a result, the three of four circles in Fig. 4.4(a), corresponding to the scattering between states with opposite spins, are suppressed. Finally, a circle corresponding to the scattering between the states with the same spin remains.

Lastly, consider a single circular CEC with chiral spin as shown in Fig. 4.4(c)^{*2}. In this case, all the backscattering is forbidden, resulting in the absence of a clear pattern in QPI.

Because the QPI pattern does not directly reflect the CEC, the interpretation of the data needs careful treatment. Conversely, the QPI pattern includes various characteristics of electrons in the crystal; careful analysis should help to understand the characteristics of the materials.

4.3 Analysis of QPI

The square of the wave function (as in eq. 4.4) is directly related to the density of electron or DOS. Thus, the standing wave can be observed by STM, in which the tunneling current depends on the DOS. By using the lock-in technique, the spatial distribution of the DOS at specific energy can be measured, in which the standing waves appear. This measurement is called STS mapping or dI/dV mapping. Figure 4.5 shows the bias-dependent change of the dI/dV mapping on Au(111). The wavy patterns are the standing wave, which originates from the QPI, and the bias-dependent evolution of the standing wave is clear. The wavenumber of the standing wave can be visualized by Fourier transform of the image, as shown in Fig. 4.5(d)-(f), which we call QPI pattern. The QPI pattern consists of a circle, reflecting the free-electron like surface band on Au(111). We can express the surface band as

$$E = \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2) + E_0. \quad (4.6)$$

The shape of the band is characterized by the effective mass (m^*) and the bottom energy (E_0), which can be determined by analyzing the evolution of the QPI pattern. If we assume a simple circular CEC (as is the case in Fig. 4.2), the radius of the circle in the QPI pattern corresponds to the twice of the radius

^{*2} This situation can be realized on the surface of topological insulator

Table 4.1 QPI analysis

E(meV)	q(2π/nm)	k(2π /nm)
200	3.80	1.90
-20	1.30	1.65
-200	2.60	1.30
-300	1.96	0.98
-400	1.11	0.56

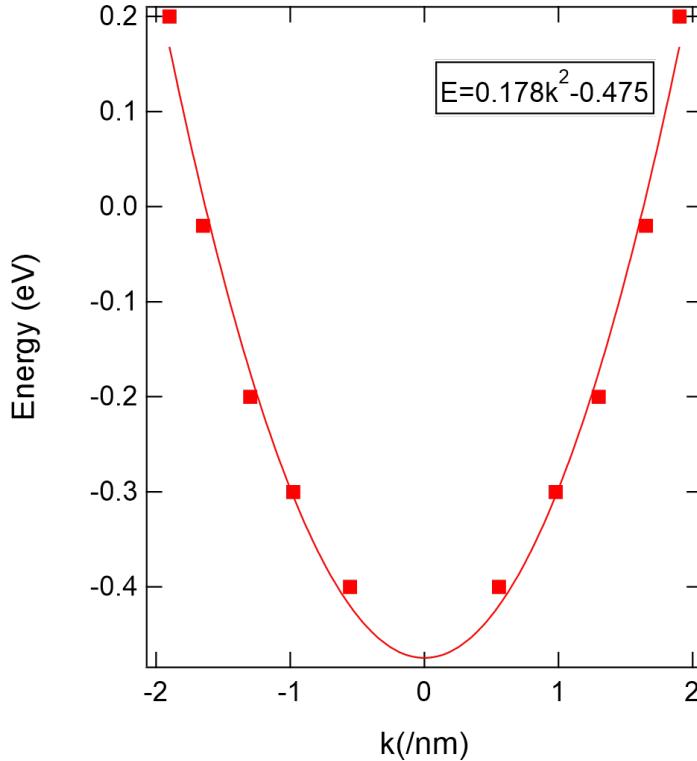


Fig. 4.6 Plot of wavenumber of standing wave, and the fitting by parabolic function.

of the circle in CEC ($|q| = 2|k|$). Table 4.1 summarizes the q values in the QPI pattern and derived k values.

Figure 4.6 shows the plot of k at each energy with the fitting curve,

$$E[eV] = 0.178k^2[nm^{-1}] - 0.475[eV]. \quad (4.7)$$

The fitting indicates that the bottom of the parabola is located at -0.475 eV. The coefficient of the first term corresponds to the $\frac{\hbar^2}{2m^*}$, so the effective mass can also be calculated from the fitted value. Note here that the equation should be converted to an MKS unit for direct comparison. After the proper conversion, the fitting equation is rewritten as

$$E[J] = 2.85 \times 10^{-38}k^2[m^{-1}] - 0.76 \times 10^{-19}[J]. \quad (4.8)$$

Therefore, the effective mass is calculated to be

$$\frac{\hbar^2}{2m^*} = 2.85 \times 10^{-38} \quad (4.9)$$

$$\therefore m^* = 1.93 \times 10^{-31}[kg] \quad (4.10)$$

Comparing with the mass of electron $m_e = 9.11 \times 10^{-31}$,

$$\frac{m^*}{m_e} = 0.21. \quad (4.11)$$

Now we got the detailed characteristics of the band structure of the surface state of Au(111): the surface state starts at -0.475eV , and the effective mass is 0.21. These values are comparable with the photoemission experiments, which showed $E_0 = -408 \text{ meV}$ and $m^* = 0.284m_e$ [16].

Chapter 5

Appendix

5.1 Quantum tunneling of an electron between tip and sample

Quantum tunneling of an electron between a tip and a sample can be simplified to a one-dimensional model, as shown in Fig. 5.1. An incident electron feels a potential of V_0 between $x = 0$ to d . According to quantum mechanics, even if the kinetic energy of the electron is lower than the potential, the electron can penetrate the potential wall with a certain possibility. This phenomenon is called as quantum tunneling effect. Schrödinger equations of each region are described as follows:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_1 = E\psi_1 \quad (x < 0) \quad (5.1)$$

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V_0 \right) \psi_2 = E\psi_2 \quad (0 \leq x \leq d) \quad (5.2)$$

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi_3 = E\psi_3 \quad (d < x) \quad (5.3)$$

Solutions of these equations are expressed as follows^{*1}

$$\psi_1 = A_1 \exp(ik_1 x) + A_2 \exp(-ik_1 x) \quad (5.4)$$

$$\psi_2 = B_1 \exp(k_2 x) + B_2 \exp(-k_2 x) \quad (5.5)$$

$$\psi_3 = C_1 \exp(ik_1 x) \quad (5.6)$$

Here,

$$k_1 = \frac{\sqrt{2mE}}{\hbar}, \quad k_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \quad (5.7)$$

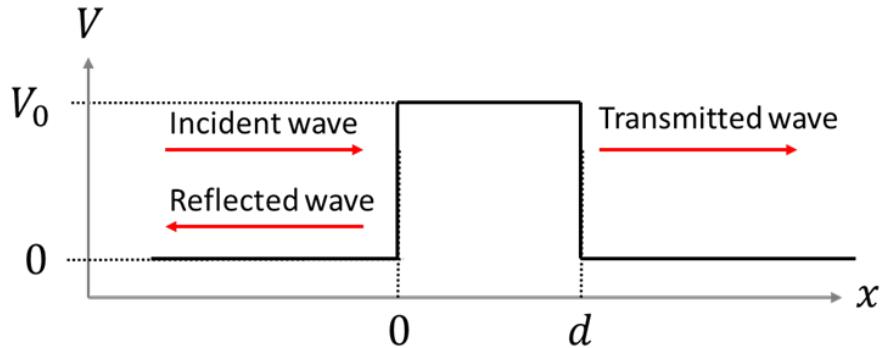


Fig. 5.1 One-dimensional model of quantum tunneling. V is the potential energy, and x is the position coordinate. An electron incident to the potential region at $x=0$. The electron is stochastically reflected or transmitted.

^{*1} In case of $E < V_0$.

The following relations are derived to satisfy boundary conditions^{*2}:

$$A_1 + A_2 = B_1 + B_2 \quad (5.8)$$

$$ik_1(A_1 - A_2) = k_2(B_1 - B_2) \quad (5.9)$$

$$B_1 \exp(k_2 d) + B_2 \exp(-k_2 d) = C_1 \exp(ik_1 d) \quad (5.10)$$

$$k_2 \{B_1 \exp(k_2 d) - B_2 \exp(-k_2 d)\} = ik_1 C_1 \exp(ik_1 d) \quad (5.11)$$

The probability of electron tunneling is expressed by the intensity of the incident wave and the transmitted wave as follows:

$$T = \left| \frac{C_1}{A_1} \right| = \frac{1}{1 + \frac{(k_1^2 + k_2^2)^2}{4k_1^2 k_2^2} \sinh^2(k_2 d)} \quad (5.12)$$

If $k_2 d \gg 1$, $\sinh(k_2 d) \sim \exp(k_2 d)/2$ and the first term of the denominator can be ignored. Therefore,

$$T = \frac{16k_1^2 k_2^2}{(k_1^2 + k_2^2)^2} \exp(-2k_2 d) \propto \exp(-2k_2 d) \quad (5.13)$$

As a result, we obtained the tunneling probability of an electron through a potential barrier.

Then, we will consider the quantum tunneling through non-constant potential. A non-constant potential can be regarded as a successive rectangular potential, as shown in Fig. 5.2. Considering that the electron passes through successive rectangular potentials with the width of Δx , the tunneling probability is expressed as follows:

$$\begin{aligned} T &= \exp[-2k_2(x_1)\Delta x] \times \exp[-2k_2(x_1 + \Delta x)\Delta x] \times \cdots \times \exp[-2k_2(x_2)\Delta x] \\ &= \exp \left[-2 \int_{x_1}^{x_2} k_2(x) dx \right] \end{aligned} \quad (5.14)$$

The potential barrier during the STM operation is deduced from the energy diagram as shown in Fig. 5.3. Bias voltage V is applied between tip and sample, which yields the energy difference of eV . Electrons in the tip with energy between $E_{F,T}$ and $E_{F,T} - eV$ can transmit to sample through the vacuum. The potential energy between the tip and sample is expressed by the work functions, bias voltage, and tip-sample distance (d). The potential energy is assumed to change with x constantly to simplify the situation. Considering that $E_{F,S} = 0$, the tunneling probability of an electron with an energy of E is expressed as follows:

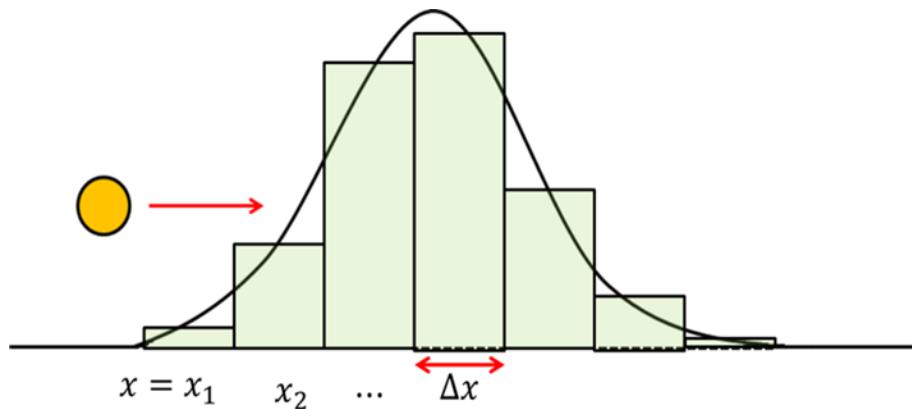


Fig. 5.2 A non-constant potential. The potential can be divided into successive rectangular potential walls.

^{*2} The value and slope (derivative) should match at the boundary.

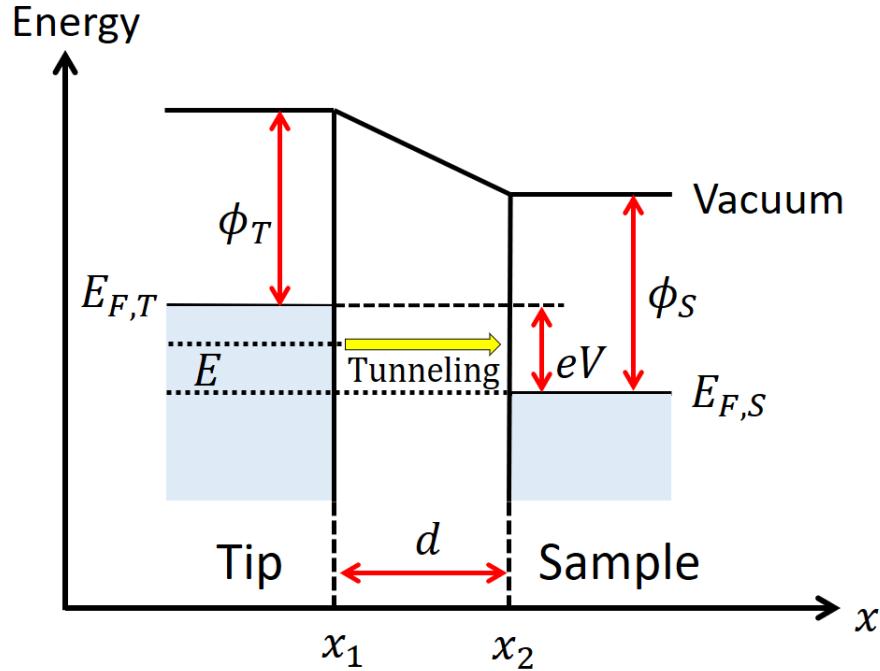


Fig. 5.3 Energy diagram during the STM operation. $E_{F,T}$ and $E_{F,S}$ are the Fermi energy of the tip and sample, respectively. Bias voltage V is applied between tip and sample, which yields energy difference of eV . ϕ_T and ϕ_S is the work function of the tip and sample, respectively.

$$\begin{aligned} T &= \exp \left[-2 \int_{x_1}^{x_2} k_2(x) dx \right] = \exp \left[-\frac{2\sqrt{2m}}{\hbar} \int_{x_1}^{x_2} \sqrt{V(x) - E} dx \right] \\ &= \exp \left[-\frac{2d\sqrt{2m}}{\hbar} \sqrt{\phi_0 + \frac{eV}{2} - E} \right] \end{aligned} \quad (5.15)$$

Here, ϕ_0 is the average of the work functions of the tip and sample. The tunneling probability of an electron between the tip and sample is expressed as eq. 5.15. The total number of tunneling electrons is in proportion to the DOS of the tip (ρ_t) and sample (ρ_s). Therefore, the tunneling current is approximately expressed as follows:

$$\begin{aligned} I_t &= \int_0^{eV} \rho_t \rho_s T dE \\ &\sim \rho_t \rho_s T eV \\ &= I_0 \exp \left[-\frac{2d\sqrt{2m\phi_0}}{\hbar} \right] \end{aligned} \quad (5.16)$$

Here, we assumed that the DOS of both tip and sample does not depend on the energy, and the bias voltage applied between the tip and sample is significantly smaller than ϕ_0 . This equation indicates the exponential dependence of tunneling current on the distance between the tip and sample.

5.2 Calculation detail to get Equation 1.19

Here we will insert the perturbed wavefunction

$$\psi = a_\mu(t) \psi_\mu^S e^{-iE_\mu^S t/\hbar} + \sum_{\nu=1}^{\infty} c_\nu(t) \psi_\nu^T e^{-iE_\nu^T t/\hbar}. \quad (5.17)$$

into the Schrödinger equation of the whole system,

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S + U_T(t) \right) \psi. \quad (5.18)$$

Then get the projection to ψ_ν^T . Here, $U_T(t) = e^{\eta t/\hbar} U_T$. First, consider the left side.

$$i\hbar \frac{\partial \psi}{\partial t} = i\hbar \frac{\partial a_\mu(t)}{\partial t} \psi_\mu^S e^{-iE_\mu^S t/\hbar} \quad (5.19)$$

$$+ iE_\mu^S a_\mu(t) \psi_\mu^S e^{-iE_\mu^S t/\hbar} \quad (5.20)$$

$$+ i\hbar \sum_{\nu=1}^{\infty} \frac{\partial c_\nu(t)}{\partial t} \psi_\nu^T e^{-iE_\nu^T t/\hbar} \quad (5.21)$$

$$+ \sum_{\nu=1}^{\infty} E_\nu^T c_\nu(t) \psi_\nu^T e^{-iE_\nu^T t/\hbar} \quad (5.22)$$

Because we are considering the adiabatic condition, the change of the $a_\mu(t)$ is small and slow. Thus, $\frac{\partial a_\mu(t)}{\partial t} \sim 0$ and $a_\mu(t) \sim 1$. Then, consider the projection to ψ_ν^T . In other words, multiply ψ_ν^{T*} and integrate it in the whole space. Because of the assumption of the theory, all the states in eq. 5.17 is orthogonal, that are,

$$\int \psi_\mu^{T*} \psi_\nu^S d^3r \sim 0 \quad (5.23)$$

$$\int \psi_\mu^{T*} \psi_\lambda^T d^3r \sim 0 (\mu \neq \lambda) \quad (5.24)$$

$$\int \psi_\mu^{T*} \psi_\mu^T d^3r = 1. \quad (5.25)$$

By applying these assumptions, term in 5.19 and 5.20 disappear, and for term in 5.21 and 5.22, only one component remains after the projection as follows:

$$(left) = i\hbar \frac{\partial c_\nu(t)}{\partial t} e^{-iE_\nu^T t/\hbar} + E_\nu^T c_\nu(t) e^{-iE_\nu^T t/\hbar}. \quad (5.26)$$

Then, consider the right side.

$$\left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S + U_T(t) \right) \psi = \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S + e^{\eta t/\hbar} U_T(t) \right) (a_\mu(t) \psi_\mu^S e^{-iE_\mu^S t/\hbar} + \sum_{\nu=1}^{\infty} c_\nu(t) \psi_\nu^T e^{-iE_\nu^T t/\hbar}) \quad (5.27)$$

$$= \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + U_S \right) a_\mu(t) \psi_\mu^S e^{-iE_\mu^S t/\hbar} + a_\mu(t) e^{\eta t/\hbar} U_T \psi_\mu^S e^{-iE_\mu^S t/\hbar} \quad (5.28)$$

$$+ \sum_{\nu=1}^{\infty} c_\nu(t) e^{-iE_\nu^T t/\hbar} \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial z^2} + e^{\eta t/\hbar} U_T \right) \psi_\nu^T \quad (5.29)$$

$$+ \sum_{\nu=1}^{\infty} c_\nu(t) U_S \psi_\nu^T e^{-iE_\nu^T t/\hbar} \quad (5.30)$$

Consider the deformation of each term by the projection. The first term can be deformed by using the unperturbed Schrödinger equation 1.11. Thus,

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_S\right)a_\mu(t)\psi_\mu^S e^{-iE_\mu^S t/\hbar} = E_\mu^S a_\mu(t)e^{-iE_\mu^S t/\hbar}\psi_\mu^S \quad (5.31)$$

, which becomes 0 after the projection because of the orthogonality. The second term is kept after the projection as follows:

$$(second) = e^{\eta t/\hbar}e^{-iE_\mu^S t/\hbar} \int \psi_\nu^{T*} U_T \psi_\mu^S d^3r. \quad (5.32)$$

The third term can be deformed by using the unperturbed Schrödinger equation of the tip (eq. 1.10) as follows:

$$\sum_{\nu=1}^{\infty} c_\nu(t)e^{-iE_\nu^T t/\hbar} \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + e^{\eta t/\hbar}U_T\right)\psi_\nu^T = \sum_{\nu=1}^{\infty} c_\nu(t)e^{-iE_\nu^T t/\hbar}[E_\nu^T + (e^{\eta t/\hbar} - 1)U_T]\psi_\nu^T. \quad (5.33)$$

Here, $(e^{\eta t/\hbar} - 1) \rightarrow 0$ when $\eta \sim 0$. Thus, the term with $(e^{\eta t/\hbar} - 1) \rightarrow 0$ can be ignored. After the projection, the remaining is

$$(third) = E_\nu^T c_\nu(t)e^{-iE_\nu^T t/\hbar} \quad (5.34)$$

For the fourth term, all the component remains after the projection as follows:

$$(Fourth) = \sum_{\lambda=1}^{\infty} c_\lambda(t)e^{-iE_\lambda^T t/\hbar} \int \psi_\nu^{T*} U_S \psi_\lambda^T d^3r \quad (5.35)$$

Note that the sum-up number is replaced from ν to λ to distinguish from the projection state. To summarize, the right side after the projection is expressed as follows:

$$\begin{aligned} (right) &= e^{\eta t/\hbar}e^{-iE_\mu^S t/\hbar} \int \psi_\nu^{T*} U_T \psi_\mu^S d^3r + E_\nu^T c_\nu(t)e^{-iE_\nu^T t/\hbar} \\ &\quad + \sum_{\nu=1}^{\infty} c_\nu(t)e^{-iE_\nu^T t/\hbar} \int \psi_\nu^{T*} U_S \psi_\nu^T d^3r \end{aligned} \quad (5.36)$$

Finally, from eq. 5.26 and eq. 5.36, we get

$$i\hbar \frac{\partial c_\nu(t)}{\partial t} e^{-iE_\nu^T t/\hbar} = a_\mu(t)e^{\eta t/\hbar}e^{-iE_\mu^S t/\hbar} \int \psi_\nu^{T*} U_T \psi_\mu^S d^3r + \sum_{\lambda=1}^{\infty} c_\lambda(t)e^{-iE_\lambda^T t/\hbar} \int \psi_\nu^{T*} U_S \psi_\lambda^T d^3r \quad (5.37)$$

$$\therefore i\hbar \frac{dc_\nu}{dt} = \int \psi_\nu^{T*} U_T \psi_\mu^S d^3r e^{-i(E_\mu^S - E_\nu^T + i\eta)t/\hbar} + \sum_{\lambda=1}^{\infty} c_\lambda(t) \int \psi_\nu^{T*} U_S \psi_\lambda^T d^3r e^{-i(E_\lambda^T - E_\nu^T)t/\hbar}. \quad (5.38)$$

5.3 Detection of second derivative in Lock-in

In section 2.1, we used $A\sin(\omega t + \theta)$ as a reference signal to get the dI/dV signal. In this appendix, we will consider how to get the second derivative by using Lock-in detection.

The Taylor series of the modulated current is expressed as follows:

$$I(V + \Delta V) = I(V) + \frac{dI}{dV}(\Delta V \sin \omega t) + \frac{d^2 I}{dV^2} \frac{1}{2}(\Delta V \sin \omega t)^2 + \dots \quad (5.39)$$

For getting the second derivative, $A\sin(2\omega t + \theta)$ is used as a reference signal. The products are

$$\text{First term} \quad I(V) \times A\sin(2\omega t + \theta) \quad (5.40)$$

$$\begin{aligned} \text{Second term} \quad & \frac{dI}{dV}(\Delta V \sin \omega t) \times A\sin(2\omega t + \theta) \\ &= \frac{A\Delta V}{2} \frac{dI}{dV} \{\cos(\omega t + \theta) - \cos(3\omega t + \theta)\} \end{aligned} \quad (5.41)$$

$$\begin{aligned} \text{Third term} \quad & \frac{d^2 I}{dV^2} \frac{1}{2}(\Delta V \sin \omega t)^2 \times A\sin(2\omega t + \theta) \\ &= \frac{A\Delta V^2}{4} \frac{d^2 I}{dV^2} \{2\sin(2\omega t + \theta) - \sin(4\omega t + \theta) - \sin \theta\} \end{aligned} \quad (5.42)$$

After passing the LPF, the remaining signal is only the final component of the third term,

$$V_X = -\frac{A\Delta V^2}{4} \frac{d^2 I}{dV^2} \sin \theta. \quad (5.43)$$

As a result, a signal which is proportional to the second derivative is obtained. θ should be set at -90° to get the largest signal. Phase independent signal V_R can be also defined in the same way as the first derivative case.

$$V_R = \sqrt{V_X^2 + V_Y^2} = \left(\frac{A\Delta V^2}{4}\right)^2 \times \left|\frac{d^2 I}{dV^2}\right|. \quad (5.44)$$

In general, the second derivative of tunneling current can take both positive and negative values (cf. the first derivative is always the positive value except for some special cases). Thus, be careful that the V_R can change the sign of the signal.

5.4 Phase sensitivity of Lock in detection

Two kinds of outputs (R and X) are available in lock-in detection. Because of the phase sensitivity, X signal with proper phase should show lower noise. On the other hand, the advantage of the R signal is the easy measurement without the phase setting process. Here Fig. 5.4 shows the difference of the dI/dV image obtained by R and X signal. Even though the R signal shows the QPI signal on the surface, the contrast of the image is clearly improved by using the X signal with the proper phase. Because the essential feature in the image is identical, the R signal can also be used. However, remember that using X signal would result in a better S/N.

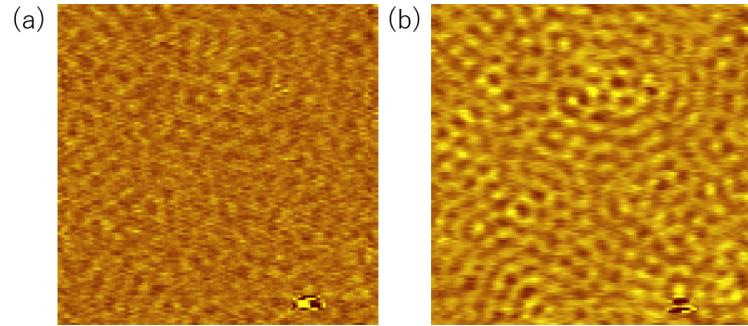


Fig. 5.4 Comparison of the mapping image with (a) R and (b) X signal in the lock-in amplifier.

5.5 Calculation of thermal drift

Consider the relationship between the square-shape STM image and parallelogram-shape scanning region with using Fig. 5.5. Define two vectors along the sides of the STM image as \vec{a} and \vec{b} . If the size of the image is $L \times L$, the two vectors are expressed as

$$\vec{a} = \begin{pmatrix} L \\ 0 \end{pmatrix}, \vec{b} = \begin{pmatrix} 0 \\ L \end{pmatrix} \quad (5.45)$$

The vectors along the parallelogram (\vec{a}' and \vec{b}') are expressed as follows:

$$\vec{a}' = \begin{pmatrix} L + v/2N \\ w/2N \end{pmatrix}, \vec{b}' = \begin{pmatrix} v \\ L + w \end{pmatrix} \quad (5.46)$$

The coefficient $1/2N$ originates from the time for scanning one line. In general setting, the STM tip traces a line twice (forward and backward). Then, the time for scanning from left to right (forward scan) is $1/2N$ of total scanning time. The transformation matrix is expressed as follows^{*3}:

$$A = \frac{1}{L} \begin{pmatrix} L + v/2N & v \\ w/2N & L + w \end{pmatrix} \quad (5.47)$$

By applying this transformation matrix to the STM image, image without distortion is obtained. The values of thermal drift (v and w) are necessary for the transformation. These values can be calculated from the apparent drift in STM images which are obtained consecutively. Figure 5.5(a) shows the relationship of the scanning regions of two consecutive STM images. The size of the thermal drift is the same as the change of the start position of two scannings. The start position of the scanning can be calculated from the STM images, as shown in Fig. 5.5(b). The vectors \vec{d}_1 and \vec{d}_2 are the apparent position of a defect in the STM images. These vectors are converted to the real scale by applying the transformation matrix:

$$\vec{d}'_1 = A\vec{d}_1 \quad (5.48)$$

$$\vec{d}'_2 = A\vec{d}_2 \quad (5.49)$$

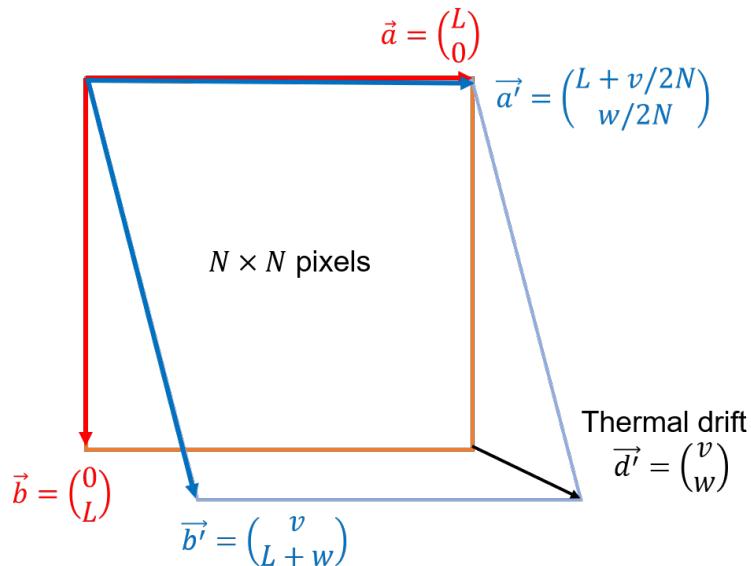


Fig. 5.5 The relationship between the STM image and scanning region.

^{*3} Confirm that $\vec{a}' = A\vec{a}$ and $\vec{b}' = A\vec{b}$

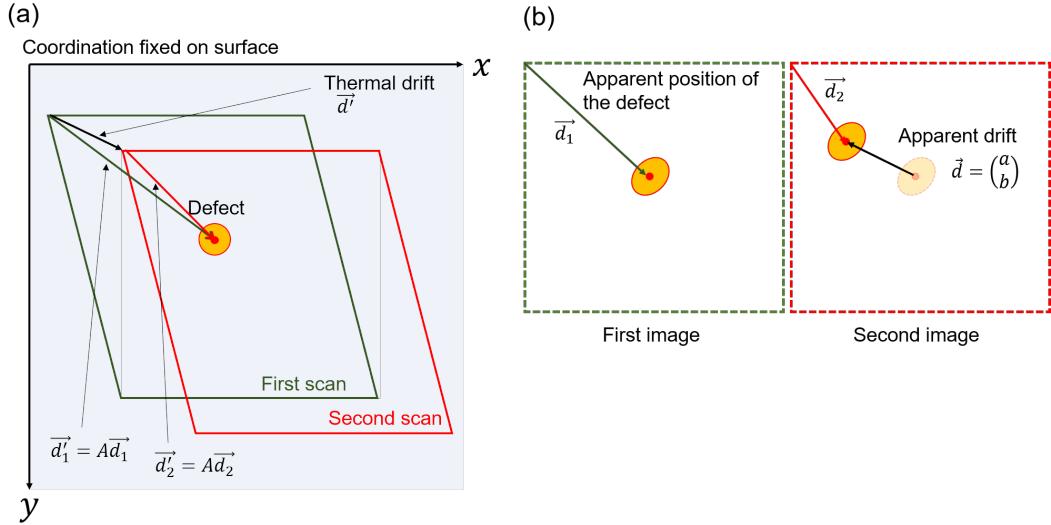


Fig. 5.6 (a) The relationship between the scanning regions, the position of defect on the surface, and thermal drift. (b) The vectors that are obtained from the STM images.

The thermal drift vector is given by

$$\begin{aligned}\vec{d}' &= \vec{d}'_1 - \vec{d}'_2 \\ &= A(\vec{d}_1 - \vec{d}_2) \\ &= -A\vec{d}\end{aligned}\tag{5.50}$$

$$\therefore \begin{pmatrix} v \\ w \end{pmatrix} = -\frac{1}{L} \begin{pmatrix} L + v/2N & v \\ w/2N & L + w \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}\tag{5.51}$$

This equation indicates that the thermal drift is obtained by the direct transformation of the apparent drift. The values (v, w) are obtained by solving eq. 5.51,

$$v = -\frac{a}{1 + a/2NL + b/L}\tag{5.52}$$

$$w = -\frac{b}{1 + a/2NL + b/L}\tag{5.53}$$

The specific values of the transformation matrix is obtained by substituting v and w with using eq. 5.52 and 5.53. The image without distortion is obtained by applying the transformation matrix to the image.

5.6 Experiential error in STM observation (STM 2)

One of the dominant error sources in STM observation is the thermal drift, which distorts the image. Because the thermal drift is induced by the change of the temperature of the sample, it would be suppressed after the long cooling time. Therefore, we checked how the thermal drift is suppressed by time, as shown in Fig. 5.7. The thermal drift gradually attenuates and finally becomes less than 0.01 nm/min. Considering that the sample is cooled for 4 hours before starting the experiment, it takes over 8 hours until the thermal drift reaches the smallest value. Or, if the approaching process induces the temperature change, the time to getting the smallest thermal drift is about 5 hours after finish approaching, as we started the scanning 30 minutes after finish approaching.

Creeping also distorts the image, as explained in section 3.2. We checked how long it is necessary to remove the creeping after moving the tip. It is evaluated by continuous imaging after moving the tip from the center to $(x, y) = (-350\text{nm}, 0)$. The STM image in Fig. 5.8 is the first image after the moving. Nonlinear distortion is observed especially on the top of the image, which is caused by the creeping. The creeping effect decays as shown in the center of Fig. 5.8, which indicates that the creeping becomes less than 0.02 nm/min after 40 minutes of scanning.

The scanning speed and scanning region slightly affect the images. Figure 5.9 shows how the scanning speed affects the apparent lattice constant of Si(111)-7 \times 7. The faster scanning speed overestimates the lattice constant^{*4}. It would be because the piezoelectric can not follow the voltage change (as the same effect as the creeping) if the scanning is too fast. Therefore, the actual scanning region is smaller than the set value. Although the scanning region becomes close to the intended value by using the slow scanning speed, the thermal drift and creeping become large in the slow scanning condition.

Figure 5.10 shows the dependence on the scanning size. The gray and orange dots represent the lattice

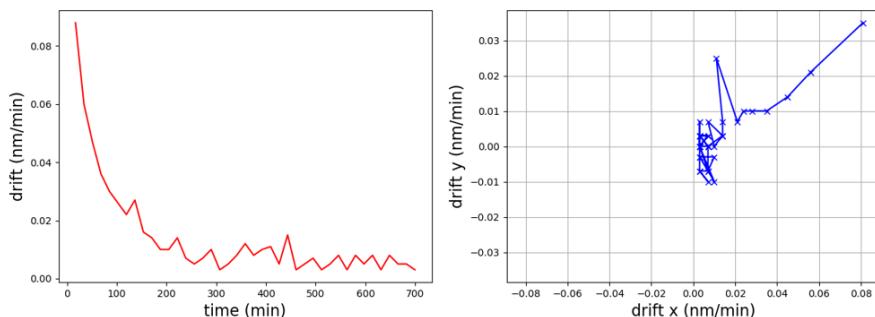


Fig. 5.7 (Left) Time dependent change of the thermal drift (Right) time dependent change of the drift direction.

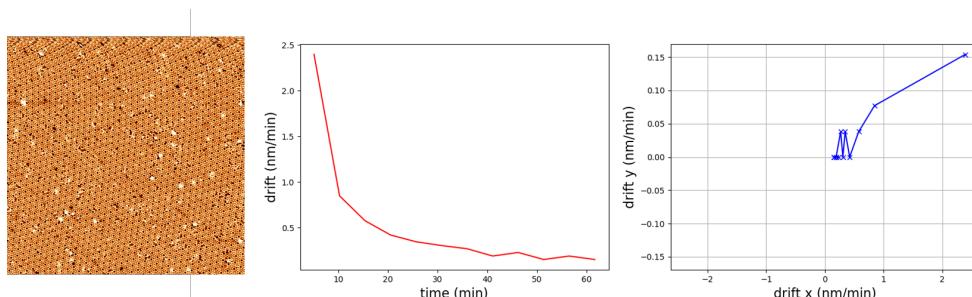


Fig. 5.8 (Left) STM image with distortion caused by creeping (Center) time dependent change of creeping speed (Right) time dependent change of the creeping direction.

^{*4} Note that the calibration in this observation is not correct (the lattice constant of Si(111) is 2.68 nm). Just look at the change of the value.

constant in three directions with 20 and 100 nm scanning sizes, respectively, and the blue dots are the averaged values. While the average values are almost identical, the lattice constant variation is larger in small scanning size because the drift largely affects small scanning size.

The scanning region also affects the image size, as shown in Fig. 5.11. The size of the lattice decreases as the scanning region become far from the center. The bottom two figures show how the scanning region affects the lattice constant in three different directions. When the scanning is performed at the region with a large X, the lattice constant 3 is strongly affected, while 1 is not. However, the tendency is the opposite when the scanning is performed at the large Y.

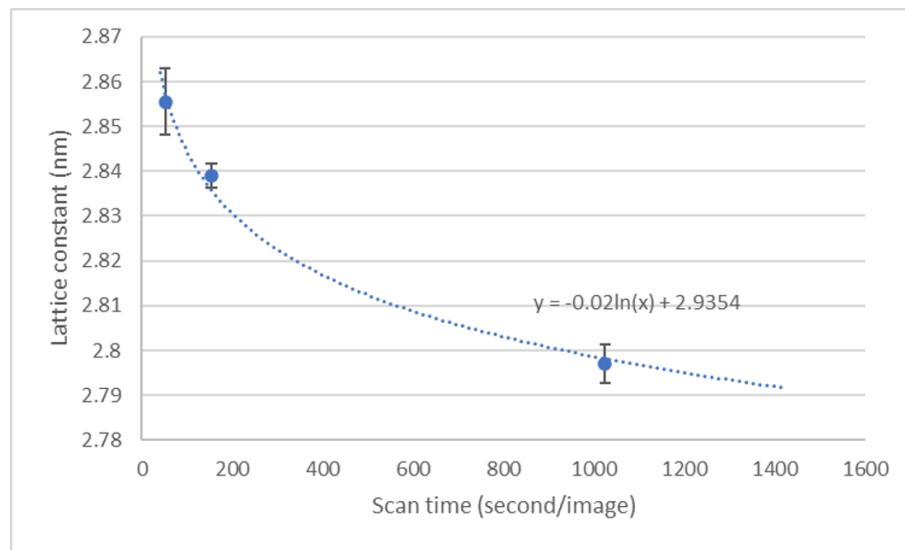


Fig. 5.9 The effect of scanning speed in the size of the unit cell.

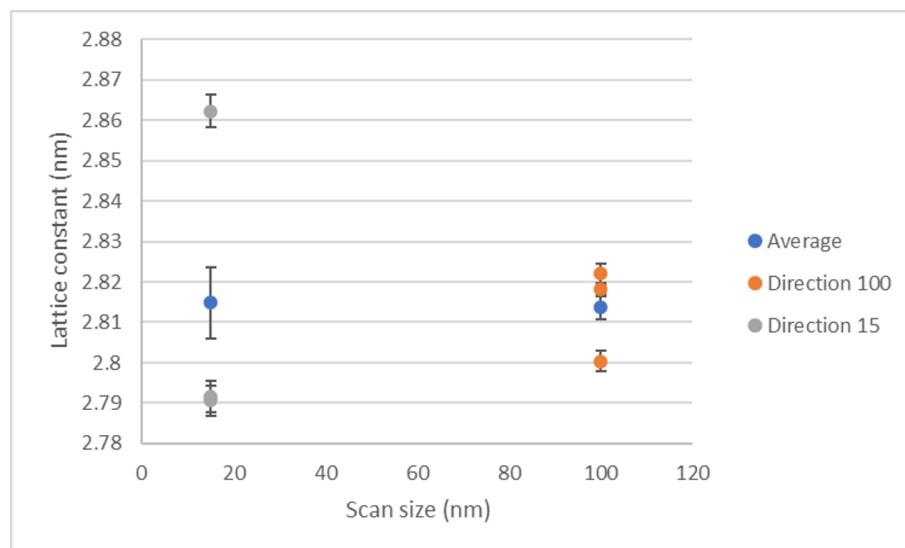


Fig. 5.10 The effect of scanning size in the size of the unit cell.

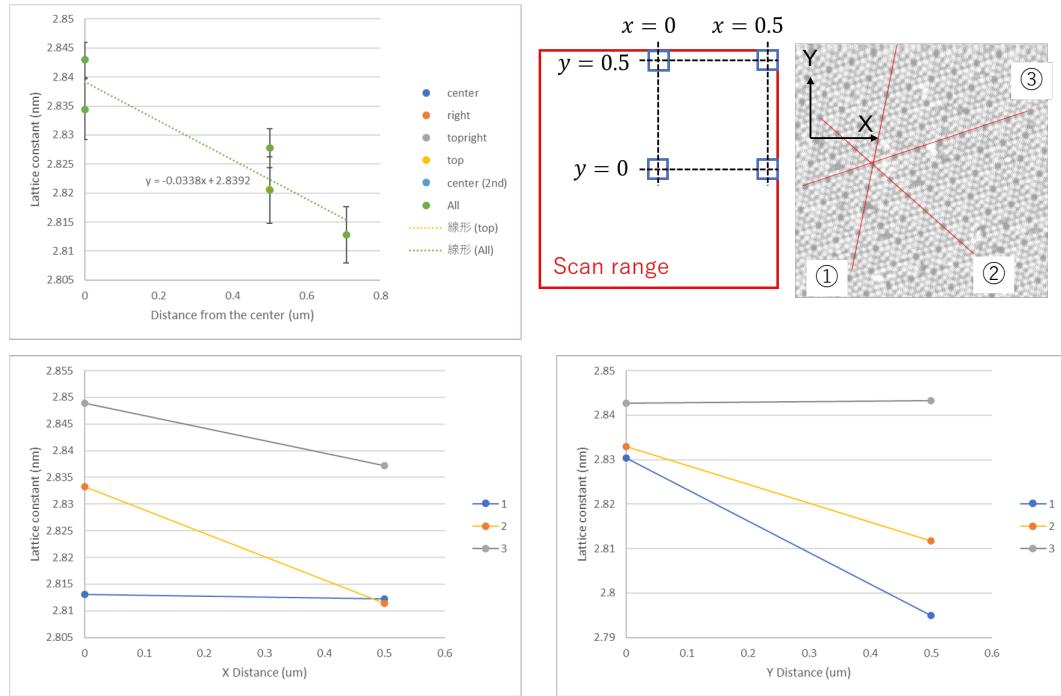


Fig. 5.11 The effect of scanning region in the size of the unit cell.

5.7 Affine calibration

Not yet

Bibliography

- [1] G. Binning, H. Rohrer, Ch. Gerber, and E. Weibel, "Surface studies by scanning tunneling microscopy", Phys. Rev. Lett. 49, 57 (1982).
- [2] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, "7 x 7 reconstruction on Si(111) resolved in real space", Phys. Rev. Lett. 50, 120 (1983).
- [3] J. A. Stroscio, R. M. Feenstra, and A. P. Fein, "Electronic Structure of the Si(111) 2 × 1 Surface by Scanning-Tunneling Microscopy", Phys. Rev. Lett. 57, 2579 (1986).
- [4] S. DeVore, A. Gauthier, J. Levy, and C. Singh, "Improving student understanding of lock-in amplifiers", Am. J. Phys. 84, 52 (2016).
- [5] Standard Research systems, "About Lock-in Amplifiers", Retrieved April 8, 2020, from <https://www.thinksrs.com/downloads/pdfs/applicationnotes/AboutLIAs.pdf>
- [6] Zurich Instruments, "Principles of lock-in detection and the state of the art", Retrieved April 8, 2020, from https://www.zhinst.com/sites/default/files/li_primer/zi_whitepaper_principles_of_lock-in_detection.pdf
- [7] Standard Research Systems, "MODEL SR 830 DSP Lock-In Amplifier", Retrieved April 8, 2020, from <https://www.thinksrs.com/downloads/pdfs/manuals/SR830m.pdf>
- [8] M. Morgenstern, D. Haude, Chr. Meyer, and R. Wiesendanger, "Experimental evidence for edge-like states in three-dimensional electron systems", Phys. Rev. B 64, 205104 (2001).
- [9] K. Sagisaka (2018), "Scanning Tunneling Spectroscopy". In The Surface Science Society of Japan (Ed.), "Compendium of Surface and Interface Analysis", Springer Singapore, pp. 607.
- [10] R. J. Hamers, "Atomic-Resolution Surface Spectroscopy with the Scanning Tunneling Microscope", Annu. Rev. Phys. Chem. 40, 531 (1989).
- [11] W. Krenner, D. Kühne, F. Klappenberger, and J. V. Barth, "Assessment of Scanning Tunneling Spectroscopy Modes Inspecting Electron Confinement in Surface-Confining Supramolecular Networks", Sci. Rep. 3, 1454 (2013).
- [12] S. Lounis, "Theory of Scanning Tunneling Microscopy", arXiv:1404.0961v1.
- [13] J. Bardeen, "Tunneling from a Many-Particle Point of View", Phys. Rev. Lett. 6, 57 (1961).
- [14] J. Tersoff, and D. R. Hamann, "Theory and Application for the Scanning Tunneling Microscope", Phys. Rev. Lett. 50, 1998 (1983).
- [15] J. Tersoff, and D. R. Hamann, "Theory of scanning tunneling microscope", Phys. Rev. B 31, 805 (1985).
- [16] S. D. Kevan, and R. H. Gaylord, "High-resolution photoemission study of the electronic structure of the noble-metal (111) surfaces", Phys. Rev. B 36, 5809 (1987).