Single- and many-particle description of scanning tunneling spectroscopy

Mikko M. Ervasti\textsuperscript{a}, Fabian Schulz\textsuperscript{b}, Peter Liljeroth\textsuperscript{b}, Ari Harju\textsuperscript{a,\ast}

\textsuperscript{a}COMP Centre of Excellence, Department of Applied Physics, Aalto University School of Science, PO Box 11100, 00076 Aalto, Finland
\textsuperscript{b}Department of Applied Physics, Aalto University School of Science, PO Box 15100, 00076 Aalto, Finland

Abstract

Scanning tunneling spectroscopy measures how a single electron with definite energy propagates between a sample surface and the tip of a scanning tunneling microscope. In the simplest description, the differential conductance measured is interpreted as the local density of states of the sample at the tip position. This picture, however, is insufficient in some cases, since especially smaller molecules weakly coupled with the substrate tend to have strong Coulomb interactions when an electron is inserted or removed at the molecule. We present theoretical approaches to go from the non-interacting and single-particle picture to the correlated many-body regime. The methodology is used to understand recent experiments on finite armchair graphene nanoribbons and phthalocyanines. We also theoretically discuss the strongly-correlated model system of fractional quantum Hall droplets.

Keywords: Scanning tunneling spectroscopy, Many-body theory, Graphene nanoribbons, Phthalocyanines

1. Introduction

Scanning tunneling spectroscopy (STS) combines atomic spatial resolution with current-voltage spectroscopy, making it possible to probe the electronic structure of single molecules and atoms \cite{1-8}. Such experiments are of fundamental importance for understanding charge and spin transport in nanoscale systems. STS is based on the scanning tunneling microscope (STM), an instrument that measures the quantum-mechanical tunneling current between a conducting sample and a sharp probe tip to generate a topographic image of the sample surface. As will be discussed in detail in this review, the tunneling current depends on the overlap between the tip and sample wave functions, and measurement of the tunneling conductance as a function of applied bias voltage allows deducing the local density of states (LDOS) of the sample.

Measurements on molecular systems are usually interpreted in terms of single-particle molecular orbitals \cite{9-11}. However, while optical spectroscopy probes the excitation spectrum of a molecule with a fixed number of electrons, STS measures the spectrum in case an electron is inserted or removed, due to tunneling between the tip and molecule. If the molecule is small enough, the charging energy to overcome the Coulomb repulsion of the electrons of the neutral molecule can be substantial. Therefore, to describe the eigenstates and -energies of especially the charged molecule properly, it is not clear a priori whether the single-particle interpretation suffices. Even so, many-body effects, which have been studied in detail in correlated electron systems such as (high Tc) superconductors \cite{12,13}, are usually neglected. The more general quantity that is applicable beyond the single-particle description, and one that corresponds to LDOS and the STS $dI/dV$ measurements, is the single-particle spectral function \cite{14}. It will be discussed extensively in the current review.

In order to compare to the experimental measurements, one can do theory and calculations in various levels of sophistication. We will discuss mainly the three cases, namely the non-interacting, the single-particle (or mean-field or independent particles), and the interacting many-body picture. The various pictures can be defined by stating that in the first two the states are written as single Slater deter-
minants, namely assuring that the particles occupy well defined molecular orbitals, whereas in the interacting case the states can be strongly correlated. In the non-interacting picture, the energies can be written as sums of occupied molecular orbital energies, whereas in the other two pictures the energies can be evaluated in principle using the full interacting Hamiltonian. We compare the theory that results in various pictures and present examples of systems where a more sophisticated description is needed. In the experimental side, it is interesting to find molecular systems that have visible features that can be explained only in the many-body picture. In the theory side, it is important to understand when a proper treatment of correlations and methods that go beyond the mean-field approaches, such as the density-functional theory (DFT), are required.

Furthermore, STS can potentially measure also higher-order tunneling processes between the tip, molecule and substrate, which can result in coherent many-body transport phenomena, such as cotunneling and the Kondo effect. For instance, in the Kondo effect a many-body singlet ground state forms between an unpaired spin in a molecule and a screening cloud of quasi-free electrons. However, such processes go beyond the description of the states on the molecular sample alone, and they are out of scope in the current review, as we will solely focus on observing intrinsic many-body effects on molecular systems using scanning tunneling spectroscopy.

This Review is organized as follows. Section 2 gives a soft introduction to scanning tunneling spectroscopy. The spectral function is discussed in various levels of theory in Section 3. This consists of the non-interacting, single-particle, and interacting descriptions of the sample. Section 4 discusses the effect of the substrate on the STS measurements, comparing especially the cases of a thin insulating substrate layer to a fully metallic substrate. Metallic substrates tend to bring the molecule ionization (HOMO) and affinity (LUMO) levels closer together. Section 5 introduces example systems of finite armchair graphene nanoribbons, phthalocyanines, and strongly-correlated fractional quantum Hall droplets on various substrates. Finally, Section 6 concludes.

2. Basic principles of scanning tunneling spectroscopy

STM is based on measuring the tunneling current flowing between an atomically sharp probe ("tip") and a conductive sample. In the simple approximation of a metal-vacuum-metal tunneling junction, and assuming the work function \( \phi \) of tip and sample are equal, the current \( I \) across the barrier is given by

\[
I \propto e^{-2\kappa z} \quad \text{with} \quad \kappa = \sqrt{2m \phi / \hbar},
\]

where \( z \) is the tip-sample distance, \( \hbar \) is Planck’s constant divided by \( 2\pi \), and \( m \) the mass of the electron. For typical values of metallic work functions of around 5 eV, the current will decrease by one order of magnitude if the tip-sample distance is increased by 1 Å. Due to this exponential decay, essentially all of the electrons tunnel between the last atom of the tip and the substrate. This is the reason why STM can be used to obtain atomically resolved images of the sample surface.

STM imaging is typically carried out in the constant-current mode, where a feedback loop is used to adjust the \( z \)-position of the tip in order to maintain a constant tunneling current. Recording the \( z \)-signal yields a topography image of the sample surface. While this is related to the actual topography of the sample, it is also influenced by electronic effects as discussed below. Alternatively, the experiment can be carried out in constant-height mode by keeping the height of the tip constant and recording the changes in the tunneling current. This mode is useful on atomically flat samples and in the case of examining flat molecules on well-defined surfaces.

In addition to the tip-sample distance, the tunneling current depends on the local density of electronic states (LDOS) \( \rho_{\text{LDOS}} \) close to the Fermi level of the sample. The sample LDOS can be probed in great detail by tunneling spectroscopy, i.e. by measuring the current as a function of the bias voltage. Understanding tunneling spectroscopy usually starts with the Bardeen picture of the tunneling process, where the current can be expressed in terms of the tunneling matrix element \( M_{\mu \nu} \) between states \( \psi_{\mu} \) in the tip and \( \psi_{\nu} \) in the sample:

\[
I = \frac{2\pi e}{h^2} \sum_{\mu, \nu} f(E_{\mu})[1 - f(E_{\nu} + eV)]
\times |M_{\mu \nu}|^2 \delta(E_{\mu} - E_{\nu})
\]
where \( f(E) \) is the Fermi function, \( V \) the applied bias between the sample and the tip, and \( E_\mu \) and \( E_\nu \) are the energies of states \( \mu \) and \( \nu \). The tunneling matrix element is given by

\[
M_{\mu\nu} = \frac{\hbar^2}{2m} \int_S (\psi^*_\mu \nabla \psi_\nu - \psi^*_\nu \nabla \psi_\mu) \cdot dS,
\]

(3)

where the surface \( S \) can be any surface lying within the barrier. This theory can be adapted to modeling STM as shown by Tersoff and Hamann \[20, 21\]. They expressed the surface wave function \( \psi_\nu \) in a very general form as a sum over Bloch waves in the plane of the surface, decaying exponentially into the vacuum. The tip was treated by approximating it as a locally spherical potential well while neglecting its detailed atomic structure. In the barrier region, the tip wave function \( \psi_\mu \) thus exhibits an asymptotic spherical form around the center of curvature of the tip \( r_0 \). This yields as the matrix element

\[
M_{\mu\nu} \propto \psi_\nu(r_0).
\]

(4)

This is known as the s-wave tip model. This matrix element is modified if the higher angular momentum orbitals of the tip are important \[2\]. For example, for a \( p_z \)-wave tip

\[
M_{\mu\nu} \propto \frac{\partial \psi_\nu(r_0)}{\partial E}.
\]

(5)

The current for finite bias voltage and temperature is the convolution of the tip and surface DOS integrated over energy (converting the sum in Eq. 2 into an integral) \[2\]

\[
I = \frac{4\pi e}{\hbar} \int_{-\infty}^{\infty} [f(E_F - eV + \epsilon) - f(E_F + \epsilon)] \rho_s(r_0, E_F - eV + \epsilon) \rho_t(E_F + \epsilon) T(\epsilon, V) d\epsilon,
\]

(6)

where \( \rho_t(E) \) is the tip density of states, transmission \( T(E, V) \) contains the energy dependence of the matrix elements and

\[
\rho_s(r_0, E) = \sum_\nu |\psi_\nu(r_0)|^2 \delta(E - E_\nu).
\]

(7)

is the local density of states (LDOS) probed at the position of the tip. At low temperatures, and if \( \rho_t \) and \( T \) do not depend (strongly) on energy, the differential conductance \( dI/dV \) is proportional to the sample DOS at voltage \( V \):

\[
dI/dV \propto \rho_s(E_F + eV)
\]

(8)

This very important result forms the basis for all scanning tunneling spectroscopy (STS) experiments, where the \( dI/dV \) signal is interpreted as the LDOS of the sample that can be probed with atomic resolution. While different procedures for recovering the LDOS rigorously from the measured \( dI/dV \) have been discussed in the literature \[22–25\], the interpretation of \( dI/dV \propto \rho_s(E_F + eV) \) is usually sufficient. In real life, the instruments have finite energy resolution (due to finite temperature, electrical noise and other broadening mechanisms) and the LDOS should be considered as a sum of the states in the sample over the energy resolution of the instrument

\[
\rho_s(E_F + eV) \approx \sum_{\delta E} |\psi_\nu(r_0)|^2 \delta(E - E_F - \delta E).
\]

(9)

3. Spectral function in various levels of theory

In the standard description of STS, the electrons are assumed to tunnel between the tip and sample one by one. Therefore the proper quantity that simulates the \( dI/dV \) measurements is, strictly speaking, not the single-particle LDOS, but the more general single-particle spectral function \[14\]

\[
A(\nu, \omega) = A^-(\nu, \omega) + A^+(\nu, \omega),
\]

(10)

where we have separated the negative (annihilation) and positive (creation) side of the spectra to their respective functions. At zero temperature, they can be written as

\[
A^-(\nu, \omega) = \sum_n \langle \Psi_{N-1}^\nu | \hat{c}_\nu | \Psi_{N}^\nu \rangle \delta(\omega - (E_{N-1}^\nu - E_{N}^\nu - 1)),
\]

(11)

and

\[
A^+(\nu, \omega) = \sum_n \langle \Psi_{N+1}^\nu | \hat{c}_\nu | \Psi_{0}^\nu \rangle \delta(\omega - (E_{N+1}^\nu - E_{0}^\nu)),
\]

(12)

where \( |\Psi_{N}^\nu \rangle \) is the \( n \):th many-body eigenstate of the sample with \( N \) particles, \( E_{N}^\nu \) is the energy of the state \( |\Psi_{N}^\nu \rangle \), \( \omega \) is the energy that corresponds to the bias voltage in STM, and \( \nu \) labels the single-particle orbital where an electron is annihilated or created. In practice, \( \nu \) often corresponds to a position eigenket \( |r \rangle \), where the tip is positioned.

The spectral function can be intuitively thought of as electrons being annihilated on the sample at
negative sample bias \( \omega \) and created at positive bias. Here, the spin is not explicitly visible in the equation, but one typically has to separately consider all cases, where either spin-up or spin-down electron is created or annihilated. The energy gap between the first creation (affinity) and annihilation (ionization) peaks (energy levels) is the fundamental gap \( \Delta = E_0^{N+1} - 2E_0^N + E_0^{N-1} \), which is also the second order central finite difference of the charging curve, namely the total energy of the system as a function of the excess charge.

3.1. Non-interacting particles: From spectral function to LDOS

We will next get familiar with this formula and show first that Eq. (9) reduces to the LDOS defined in Eq. (16) for non-interacting particles. Without interactions, \( |\Psi_0^N\rangle \) is simply a totally antisymmetric state where \( N \) lowest-energy single-particle states are occupied. In the second quantization notation, the \( i \)th single-particle state is \( |\phi_i\rangle = c_i^\dagger |0\rangle \) and it has a coordinate space wave function \( \phi_i(\mathbf{r}) = \langle \mathbf{r}|\phi_i\rangle \).

We assume this single-particle state to have energy \( \varepsilon_i \). The \( N \)-particle ground-state is then \( |\Psi_0^N\rangle = \prod_{i=1}^N c_i^\dagger |0\rangle \) and it has energy \( E_0^N = \sum_{i=1}^N \varepsilon_i \).

Next we start from the non-interacting \( N \)-particle ground-state \( |\Psi_0^N\rangle \), and we would like to obtain the rate for one of the particles to tunnel out from the system. The spectral density to tunnel out from a single-particle state \( |\nu \rangle \) is given by

\[
\sum_n \left| \langle \Psi_0^{N-1}|c_\nu|\Psi_0^N\rangle \right|^2 \delta(\omega - (E_0^N - E_0^{N-1})) ,
\]

where state \( |\Psi_0^{N-1}\rangle \) has one less particle. The non-interacting eigenstates of \( N - 1 \) particles are simply given by configurations \( \prod_{i=1}^{N-1} c_i^\dagger |0\rangle \), where the occupied orbitals are given by a set of indices \( \{\eta_i\}_{i=1}^{N-1} \), different for each state \( |\Psi_0^{N-1}\rangle \). To get a non-zero contribution for \( \left| \langle \Psi_0^{N-1}|c_\nu|\Psi_0^N\rangle \right|^2 \), the configurations \( |\Psi_0^{N-1}\rangle \) and \( |\Psi_0^N\rangle \) can only differ by one occupation. This means that the occupations of \( |\Psi_0^{N-1}\rangle \) have to be \( N - 1 \) indices between one and \( N \) (and because of fermions, the occupations can not be the same for two particles) and \( \nu \) has to be the index of the remaining occupied state. In other words, the \( N \) configurations giving non-zero contribution are simply \( c_\nu^\dagger |\Psi_0^N\rangle \), with \( 1 \leq \nu \leq N \). For each fixed \( \nu \) that corresponds to an occupied state in \( |\Psi_0^N\rangle \), the overlap is one, meaning that all the tunneling peaks have the same height. The peak positions are given by \( \delta(\omega - (E_0^N - E_0^{N-1})) \).

Because the particles are here non-interacting, the total energy is a sum of the occupied single-particle state energies. The energy of the state \( |\Psi_0^{N-1}\rangle \) is simple as it differs from \( |\Psi_0^N\rangle \) by occupation of single-particle state \( \nu \) only, and thus the total energy is \( E_{\text{GS}} + \varepsilon_\nu \). Therefore \( \delta(\omega - (E_0^N - E_0^{N-1})) = \delta(\omega - \varepsilon_\nu) \), where \( 1 \leq \nu \leq N \). The tunneling is thus possible at energies that correspond to the \( \nu \)th occupied single-particle energy.

If we want to find the tunneling amplitude for a specific point in space, we can again use Eq. (16). One should then consider \( \nu \) to be a state in the position basis. The operator for annihilating a particle at certain position is the field operator, and the standard notation would be \( \psi(\mathbf{r}) \) and for creation \( \psi^\dagger(\mathbf{r}) \) (see, e.g., [24]), but \( c_\nu \) and \( c_\nu^\dagger \) could also be used. One can expand the field operator in a single-particle basis as \( \hat{\psi}(\mathbf{r}) = \sum_i \phi_i(\mathbf{r}) c_i \), where \( \phi_i(\mathbf{r}) \) is the \( i \)th single-particle state. Inserting this into Eq. (16) and repeating the same calculation as above, one gets that again the tunneling at energies corresponding to the eigenstates, but now the peak height corresponding to the \( \nu \)th single-particle state is given by \( |\phi_\nu(\mathbf{r})|^2 \). This is exactly what the local density of states calculation would give:

\[
A^-(\mathbf{r}, \omega) = \sum_{\nu=1}^N |\phi_\nu(\mathbf{r})|^2 \delta(\omega - \varepsilon_\nu). \tag{12}
\]

Note that the summation in this formula is over the \( N \) occupied single-particle states.

In a similar fashion, one can consider particles tunneling into a ground state of \( N \) particles. The relevant part of spectral function is then

\[
\sum_n \left| \langle \Psi_0^{N+1}|c_\nu^\dagger|\Psi_0^N\rangle \right|^2 \delta(\omega - (E_0^{N+1} - E_0^N)) ,
\]

where the state \( |\Psi_0^{N+1}\rangle \) is a non-interacting eigenstate with \( N + 1 \) particles. The states in the sum that give non-zero contribution now are \( |\Psi_0^{N+1}\rangle = c_\nu^\dagger \Psi_{N+n}^0 \), with \( n > 0 \), meaning that the added particle occupies an originally unoccupied orbital with index \( N + n \). The total energy is again the sum of the single-particle energies and given by \( E_n = E_{\text{GS}} + \varepsilon_{N+n} \). Tunneling is thus possible for each unoccupied single-particle eigenstate with same amplitude and energy that corresponds to single-particle energies \( \varepsilon_j \) with \( j > N \).

Using the results above and expansion \( c_\nu^\dagger = \sum_i \phi_i(\mathbf{r}) c_i^\dagger \), the spectral function for tunneling in
at position $\mathbf{r}$ can be shown to be

$$A^+(\mathbf{r}, \omega) = \sum_{i=N+1}^{\infty} |\phi_i(\mathbf{r})|^2 \delta(\omega - \varepsilon_i).$$

It is interesting to note that this equation complements the spectral function of tunneling out in a sense that when both effects are combined, one gets tunneling at occupied and empty states in a single sum as

$$A(\mathbf{r}, \omega) = \sum_{i=1}^{\infty} |\phi_i(\mathbf{r})|^2 \delta(\omega - \varepsilon_i).$$

This shows that the spectral function leads to the same results as the local density of states for the non-interacting particles.

### 3.2. Mean-field description

To go beyond a simple non-interacting particle picture, we will first assume that the correlation effects are small and a mean-field approximation is sufficient. The peak positions in Eq. (2) come from the total energies of the $N$ and $N \pm 1$ particle systems, and the Coulomb energy part of these is not captured by a simple single-particle description, but approximately included by the mean-field. In doing a mean-field approximation for the many-body problem, one still considers the system to be described by a single Slater determinant, as was done for the non-interacting ground states, but one now optimizes the occupied single-particle orbitals to minimize the total energy. If we do separately the mean-field calculation for the $N$ and $N \pm 1$ particle systems, we obtain the total energies for these systems and also different single-particle orbitals in each case. The total energy differences give the tunneling peak energies. To calculate the matrix elements, it is first helpful to define the $N$-particle mean-field ground state to be $|\Psi_0^N\rangle = \hat{c}_{\eta_1}^\dagger \cdots \hat{c}_{\eta_N}^\dagger |0\rangle$, and for the $N - 1$ particle eigenstate $\hat{c}_{\mu_1}^\dagger \cdots \hat{c}_{\mu_{N-1}}^\dagger |0\rangle$. Notice that this state does not have to be the ground state. With these definitions, the matrix elements are

$$\langle \Psi_0^N | \hat{c}_p^\dagger | \Psi_{N-1}^N \rangle = \langle 0 | \hat{c}_{\eta_1}^\dagger \cdots \hat{c}_{\eta_N}^\dagger \hat{c}_{\mu_1}^\dagger \cdots \hat{c}_{\mu_{N-1}}^\dagger | 0 \rangle$$

$$= \sum_{\eta_1, \eta_2, \ldots, \eta_N} \langle \eta_1 | p \rangle \langle \eta_2 | \mu_1 \rangle \cdots \langle \eta_N | \mu_N \rangle$$

$$\times \langle 0 | \hat{c}_{\eta_1}^\dagger \cdots \hat{c}_{\eta_N}^\dagger \hat{c}_{\mu_1}^\dagger \cdots \hat{c}_{\mu_{N-1}}^\dagger | 0 \rangle$$

$$= \sum_{P \in S_N} \text{sgn}(P) \langle \eta_{P(1)} | p \rangle$$

$$\times \left[ \langle \eta_{P(2)} | \mu_1 \rangle \cdots \langle \eta_{P(N)} | \mu_{N-1} \rangle \right].$$

This means that one needs just to calculate the overlaps of the single-particle orbitals from the mean-field calculations. The other term needed is $\langle \Psi_0^N | \hat{c}_p | \Psi_{N+1}^N \rangle = \langle \Psi_{N+1}^N | \hat{c}_p^\dagger | \Psi_0^N \rangle^\dagger$, which can be readily evaluated by using the formula above.

One can also make an approximation that after the mean-field calculation is performed at particle number $N$, the same single-particle orbitals are used for the $N \pm 1$ states. This leads to a similar formula that was used for non-interacting particles, namely, one can then calculate the LDOS from the mean-field orbitals. This is actually what most calculations based on DFT do.

### 3.3. Spectral function in correlated systems

The spectral function defined in Eq. (9) can be used in the interacting many-body case that goes beyond the mean-field approximation. Any correlated many-body wave function can be written as a linear combination of the Slater determinants that are the $N$-particle configurations $|\Phi_i^N\rangle = \left( \prod_{j=1}^{N} \hat{c}_{\eta_j}^\dagger \right) |\psi_0\rangle$, where $|\psi_0\rangle = |\hat{c}_{\eta_1}^\dagger \cdots \hat{c}_{\eta_N}^\dagger |0\rangle$ are the single-particle orbitals used in the expansion. Explicitly, an arbitrary many-body state is written as

$$|\Psi_i^N\rangle = \sum_i \alpha_i^N |\Phi_i^N\rangle.$$ 

The coefficients $\{\alpha_i^N\}_{i=1}^M$ for the $n$th many-body eigenstate can be obtained by solving the many-body Schrödinger equation. The same solution also gives the many-body energies $E_i^N$. The relative energies of the many-body eigenstates determine the peak positions in the spectrum. On the other hand,
the peak weights at a position \( \mathbf{r} \) correspond to the squared amplitudes of terms like

\[
\langle \Psi^{-1} \rangle \hat{\epsilon}_r | \Psi \rangle = \frac{1}{N!} \int \prod_{i=1}^N d\mathbf{r}_i \prod_{i=1}^N \sum_{\delta(\mathbf{r} - \mathbf{r}_i)} \langle \Psi^{-1} \rangle \hat{c}_r | \mathbf{r}_1, \ldots, \mathbf{r}_N \rangle \langle \mathbf{r}_1, \ldots, \mathbf{r}_N | \Psi \rangle
\]

where \( \mathbf{r}_i \) denotes a missing coordinate. One should note that the prefactors in the formulas depend on the normalization of the coordinate representation many-body wave functions and we have followed the convention of [26]. Moreover, the terms with the creation operator can be evaluated similarly as \( |\langle \Psi^{-1} \rangle \hat{c}_r | \Psi \rangle |^2 = |\langle \Psi^{-1} \rangle \hat{c}_r | \Psi \rangle |^2 \).

The spectral function is effectively a single particle quantity such that the terms \( \langle \Psi^{-1} \rangle \hat{c}_r | \Psi \rangle \) can be expanded using single-particle orbitals. For instance, expanding using the natural orbitals \( \phi_i^N(\mathbf{r}) \) (the eigenfunctions of the one-body reduced density matrix), the natural \( (N-1) \)-states \( \phi_i^{N-1}(\mathbf{r}_2, \ldots, \mathbf{r}_N) \), and their common eigenvalues \( |c_i|^2 \) of the many-body ground state \( |\Psi\rangle \), Eq. (15) can be written as [27]

\[
\langle \Psi^{-1} \rangle \hat{c}_r | \Psi \rangle = \sum_i c_i \langle \Psi^{-1} \rangle \phi_i^{N-1}(\mathbf{r})^{\dagger} \phi_i^N(\mathbf{r}).
\]

The many-body effects on the spectral function are then more visible as the correlations of the many-body ground state that results in a wide distribution of the natural orbital occupations \( |c_i|^2 \) and the overlaps \( \langle \Psi^{-1} \rangle \phi_i^{N-1} \). If the natural orbitals are solved first, the natural \( (N-1) \)-states can be solved for instance using the identity

\[
c_i \phi_i^{N-1}(\mathbf{r}_2, \ldots, \mathbf{r}_N) = \int \prod_{i=1}^N d\mathbf{r}_i \phi_i^N(\mathbf{r})^\dagger \Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N).
\]

Furthermore, as shown above, the spectral function can be formally written as

\[
A(\nu, \omega) = \langle \mathbf{r} | \sum_i \hat{\rho}_i^N \delta(\omega - E_i) | \mathbf{r} \rangle,
\]

where \( \hat{\rho}_i^N = \sum_{\mathbf{r}j} d_i^j \phi_j^N \phi_i^N \) and the coefficients \( d_i^j \) are defined by Eq. (10), and \( E_i \) are the peak positions that depend on \( E_i^N \), \( E_i^{N+1} \) and \( E_i^{N-1} \). That is, the spectral function can be interpreted as a single-particle density matrix that is distributed in energy. In the independent particle picture the LDOS is formally of the same form, where \( \hat{\rho}_i \) correspond to pure states. The mixed state of the propagating electrons reflect the correlated many-body properties of the molecular states.

3.3.1. Monte Carlo integration

For the cases where the many-body wave function has an analytic form, the integral in Eq. (15) can still be tedious to calculate for large particle numbers. Monte Carlo integration with importance sampling is one obvious option for this integral. This is based on the fact that

\[
\langle \Psi^{-1} \rangle \hat{c}_r | \Psi \rangle = \langle \Psi^{-1} \rangle \delta(\omega - E_i) \langle \mathbf{r} | \mathbf{r} \rangle
\]

where the last equation means that the coordinates \( \mathbf{r}_2, \ldots, \mathbf{r}_N \) are sampled from distribution \( |\Psi^{-1} \rangle \phi_i^{N-1} \). The remaining coordinate \( \mathbf{r} \) specifies the point where the spectral function is evaluated at. The wave function ratio in Eq. (17) should be straightforward to calculate for analytic functions.

4. Substrate effects

If the substrate underneath couples to the sample, it can induce a noticeable effect on the sample states and energies. In such a case the full system with both the sample and the substrate has to be considered. It is possible to introduce a thin insulating layer to decouple the molecular sample from the metallic substrate. We discuss both cases in what follows.

4.1. Metallic substrate

When a molecule is brought close to a metallic substrate, its energy levels can be drastically modified. Most notably, the energy gap between the
ionization (HOMO) and affinity levels (LUMO) is reduced from that of an isolated molecule. This is due to:

1. One-body couplings between the molecule and substrate states shifting and broadening the energy levels, such as in the Fano model [29].
2. The electron-electron interactions between the molecule and the substrate perturbing [or induces changes in] the densities, which alters the molecule energy levels.
3. Electron or hole additions [or electron removals] on the molecule causing the electron states of the molecule to relax further to minimize the total energy, resulting in charge polarization (dynamical screening). The correlation energies are heavily altered, and the gap between the ionization (HOMO) and affinity levels (LUMO) substantially decreases [28]. Higher DOS at the Fermi level will result in even smaller energy gaps between the molecular energy levels [29].
4. The molecule can become polarized as well due to dynamic charge transfer from a strongly-coupled metallic substrate, especially as a result of the electron or hole additions [29].

Even with an insulating layer between the molecule and the metallic substrate, the dynamical screening by the substrate can clearly reduce the energy gaps [30]. The polarization effects can be estimated by evaluating the screened Coulomb interactions or using classical image charge corrections [28, 31]. It should be noted that the Kohn-Sham or Hartree-Fock energy levels do not take the dynamical screening of added electrons or holes into account, and such single-particle picture is insufficient in general to describe the energy peak positions of molecules on a metallic substrates.

4.2. Insulating and weakly-coupled substrates

If the molecules are electronically sufficiently decoupled from the substrate (through e.g. the use of an ultrathin insulating film), the tip-molecule-substrate system can be modeled as a double-barrier tunnel junction [1, 32–37]. In this regime, the electronic states of the molecule are not hybridized with the metallic substrate and the charge in the molecule is quantized. The tunneling is similar to single-charge tunneling in nanostructures, see e.g. [38, 39]. The resonances in the tunneling spectra correspond to (transient) addition/removal of electrons from/to the molecule. Importantly, now as the charge in the molecule is quantized, the total energies in Eq. (19) contain charging energies of the molecule. In the gas phase of the molecule, these are the ionization potential (IP) and electron affinity (EA). Of course, in the case of tunneling spectroscopy, the presence of the substrate has a significant effect on the electron-electron screening and additionally, depending on the relative alignment of the molecular resonances with the substrate Fermi level, it is also possible to observe permanent charging of the molecules [40, 41].

In the simple single-particle (constant interaction) picture, electrons will be added (at positive substrate bias) to the molecule when the bias voltage reaches the following condition (the tip Fermi level aligns with the lowest unoccupied molecular orbital, LUMO)

\[ \eta V = E_{\text{LUMO}} - E_F + \Sigma_- \]  \hspace{1cm} (18)

where \( \eta \) is the fraction of the applied bias that drops between the tip and the molecule and \( \Sigma_- \) is the polarization energy associated with electron addition into the molecule. Correspondingly, at sufficiently negative bias, hole tunneling through the highest occupied molecular orbital HOMO can occur

\[ \eta V = E_{\text{HOMO}} - E_F - \Sigma_+ \]  \hspace{1cm} (19)

where \( \Sigma_+ \) is the polarization energy associated with positive charging of the molecule (sign taken as positive). These processes correspond to an opening of a new tunneling channel and result in a step increase in the measured tunneling current, i.e. a peak in the \( dI/dV \) signal. The bias voltages corresponding to the resonances can be used to obtain the energies of the molecular orbitals w.r.t. the substrate Fermi level. However, as the resonances in \( dI/dV \) spectra correspond to temporary charging of the molecule, their energies are affected by the Coulomb energy involved in adding or removing one electron. This also implies that the STM transport gap \( \Delta V_{\text{HOMO-LUMO}} \) is not equal to the single-particle HOMO-LUMO gap (or the optical gap). They are related as follows:

\[ \eta \Delta V_{\text{HOMO-LUMO}} = E_{\text{LUMO}} - E_{\text{HOMO}} + \Sigma_- + \Sigma_+ \]  \hspace{1cm} (20)

Again, the factor \( \eta \) in Eq. (20) results from the potential distribution in the double-barrier tunnel junction and the finite bias drop between the
Figure 1: Simulated dI/dV maps of a three-monomer 7-AGNR from first principles (DFT), mean-field Hubbard model (Hubbard-MF), and the spectral function from the many-body Hubbard model (Hubbard-MB). The real-space imaging is simulated at a constant height of 3.5Å(DFT) and 4.0Åin (Hubbard-MF and Hubbard-MB). To produce the maps of the Hubbard models, each site is assumed to correspond to analytic $p_z$ orbitals with overlaps neglected. On the right, the corresponding STS spectra is shown. The many-body Hubbard model energy values are the ground state spectral function peak energies. From [42].

5. Example systems

In this section, the theory outlined above is used to model experiments on three example systems: finite graphene ribbons, phthalocyanine molecules, and fractional quantum Hall droplets.

5.1. Finite armchair graphene nanoribbons

For carbon-based materials, one would not expect strong many-body phenomena, but still the LDOS from a mean-field or density-functional-theory calculation can qualitatively differ from a many-body calculation. The main difference is that the spectral function peak energies come from the eigenenergies of the $N \pm 1$-particle systems in addition to the unperturbed $N$-particle system.

In Fig. 1 the simulated STM maps and spectra of a three-monomer-long and seven-atom-wide finite armchair graphene nanoribbon (7-AGNR) are shown. The LDOS maps from DFT and mean-field Hubbard model calculations agree well with the spectral function maps from an exact diagonalization calculation of the many-body Hubbard model. The agreement between the different computational strategies is surprisingly good. However, the spectra do not fully match, since the spectral function has a double peak structure at both sides of the Fermi level ($E = 0$). This discrepancy originates from the non-degenerate eigenenergies of the $N \pm 1$-particle systems that differ from the single-particle orbital energies of a neutral ribbon. In fact, the same double peak structure is present in the mean-field orbital energies in case of $N \pm 1$ particles. However, this is not shown in Fig. 1 as the mean-field peak positions are not calculated from the total energies but from the mean-field orbital energies of the neutral system [42]. The conclusion of this is that the correlation effects are too small to be resolved by present STM experiments, similarly to Ref. [43] for the polyaromatic molecules.

The effect of hole-doping on the finite 7-AGNRs...
is shown in Fig. 2. The neutral \((q = 0)\) AG-NRs support antiferromagnetic (AFM) or ferromagnetic (FM) ordering between the zigzag-end-localized edges, the AFM order being only slightly lower in energy. Doping the ribbons by removing one \((q = 1)\) or two \((q = 2)\) electrons quenches the magnetic order, that is resulting in spin-degenerate solution. Moreover, the molecular orbital energies are altered, as the highest occupied orbitals become unoccupied by crossing to the unoccupied side above the Fermi level. The energy gaps are also affected. Especially the energy gap between the end-localized states (HOMO-LUMO gap in the \(q = 0\) case) becomes smaller, which evidently shows as a single broadened peak in the \(q = 2\) case at around \(E = 0.35\) eV. The maps, however, show that the state ordering does not change by doping, and one can more or less understand the system by considering the single-particle picture with a changing Fermi level.

Experimental, several different types of atomically precise graphene nanoribbons have been synthesized [44]. When 7-AGNRs are deposited on Au(111) substrate, the maps with \(q = 2\) presented in Fig. 2 provide the best match with the experiments [45, 46]. The measurements show a single peak on the unoccupied side of the spectrum close to the Fermi level. The positive doping of the ribbons is consistent with the p-doping of graphene on a Au(111) substrate [47].

When the molecule is strongly coupled with a metallic substrate, the charging energy is expected to be small due to the strong screening. In addition, the charging energy depends on system size and can be negligibly small for large molecules such as graphene nanoribbons [44, 45]. Also, if the charging energy is small compared to the coupling with the substrate, the charge on the molecule is not quantized and it is possible to have molecular orbitals with non-integer occupation. The nanoribbon gaps have been analyzed in detail in Ref. [31].

When the molecule is strongly coupled with a metallic substrate, the charging energy is expected to be small due to the strong screening. In addition, the charging energy depends on system size and can be negligibly small for large molecules such as graphene nanoribbons [44, 45]. Also, if the charging energy is small compared to the coupling with the substrate, the charge on the molecule is not quantized and it is possible to have molecular orbitals with non-integer occupation. The nanoribbon gaps have been analyzed in detail in Ref. [31].

5.2. Phthalocyanine

Molecules containing heavier atoms are expected to give rise to stronger correlation effects and we will consider metal phthalocyanine molecules as an example. It has been theoretically suggested that a many-body effect can be used to tune a low-spin to high-spin transition in copper phthalocyanine (CuPc) deposited on an ultrathin insulating film under nonequilibrium conditions [35–37].

Experimental results on cobalt phthalocyanine (CoPC) deposited on an hexagonal boron nitride monolayer (hBN) on Ir(111) are shown in Fig. 3 [49]. Due to the work function modulation over the different sites of the hBN moiré pattern [50], there are two different charge states of CoPC present on the surface [51]. STS spectra for electron removal from (negative bias) and electron injection into (positive bias) neutral CoPC (CoPC\(^0\)) are shown in the up-most row in Fig. 3, while the bottom-most row shows spectra for electron removal from negatively charged CoPC (CoPC\(^-\)). One would expect that CoPC\(^-\) would simply correspond to CoPC\(^0\) through a downward shift in energy of all the observed transitions. This seems to hold true for peaks labelled A1 (at \(-0.8\) V [44], not shown here) and A2 of the anionic molecule, which could be interpreted as peaks N1 and N2 of the neutral
species shifted down by $\sim 1$ eV. However, transition A3 of CoPC$^{-1}$, which would be expected to correspond to N3 of CoPC$^{0}$, should then appear well below -1.0 V. Instead, there are two transitions A3 and A4 just below A2 at energies of $\sim$-0.3 and $\sim$-0.8 V, respectively. Note that all the transitions show satellite peaks due to electron-vibration coupling [46, 51–54].

In addition, maps of the spatial variation of the $dI/dV$ signal shown in the second and fifth row of Fig. 4 reveal that the wave functions corresponding to transitions N3 and A3 exhibit a different shape and symmetry.

While resonances N2, N3 and A2 represent transitions between ground states of differently charged CoPC, A3 and A4 correspond to transitions into N-electron excited states (electron removal from CoPC$^{-1}$), which in the single-particle picture would be obtained by simply unoccupying the single-particle states of the molecule’s ground state in their respective order, e.g. as given by a density functional theory calculation. However, this neglects the interaction of the hole with the remaining electrons, which can result in orbital reordering and mixing of multiple single-particle excitations. Such effects can be taken into account by using time-dependent density functional theory (TDDFT) [55] to explicitly calculate the many-body excited states of the molecule and project them onto a chosen set of single-particle orbitals [56]. The central rows in Fig. 4 compare the dominant contributions to the TDDFT excited states upon projection onto the Kohn-Sham eigenstates of neutral CoPC with the experimental $dI/dV$ maps, and transitions A3 and A4 can be clearly identified with the first and second many-body excited states of neutral CoPC. The experimentally observed orbital reordering is well-described by the excited states computed from TDDFT, showing that this approach allows for a detailed understanding of STS spectra beyond the prevalent single-particle picture.

5.3. Probing a fundamentally correlated state, example of $\nu = 1/3$ Laughlin state

We will end this section with a one of the best-known examples of strongly correlated state of matter, fractional quantum Hall droplet [57]. The Laughlin wave function for the $\nu = 1/3$ state is given by [58]

$$
\Psi_L(r_1 \ldots r_N) = \prod_{i<j} (z_i - z_j)^3 \prod_{i=1}^N \exp(-r_i^2/4),
$$

where $z = x + iy$ and $r^2 = x^2 + y^2$. Note that natural units for the problem are used. Now Eq. (17) can be used to calculate the spectral function for the states with Laughlin wave function. Fig. 5 shows schematically the total particle density for 30 electrons and the spectral function for creating one extra particle, changing the system to the ground state with one added electron. One can see that the bulk of the Laughlin state has a flat density but the edge has some structure. Further, the spectral function shows a structure that corresponds to a one single state in the lowest Landau level (for the function in the symmetric gauge, see e.g. [57]). One would expect that as each single-particle state in the system is 1/3 occupied on average, the tunneling would have contribution from three single-particle orbitals. This is not the case, and the spectral function is identically zero for all orbitals except one. One simple way to understand this is
based on the rotational symmetry and the resulting fact that the angular momentum of the quantum Hall droplet is a good quantum number. Thus, the tunneling electron has to have angular momentum equal to the difference of the angular momenta of the $N$ and $N+1$ particle systems. In the lowest Landau levels the orbitals are uniquely determined by the angular momentum, and thus tunneling can only happen to one single-particle orbital. In conclusion, even in such a strongly-correlated model system such as the FQH droplet, the STS is not expected to see any correlated many-body features. Experimentally, STS has been used to study Landau levels and edge states in graphene, see [59].

6. Conclusions

To conclude, we have shown how the theoretical modeling of scanning tunneling spectroscopy can be developed from the simple non-interacting particle picture to mean-field and full many-body description of the problem. For sample systems strongly contacted with a metallic substrate, the substrate can hinder the intrinsic many-body effects of the sample. Decoupling the system from the metal by an insulating layer reduces the screening effects of the substrate. Further, as the tunneling to metal is reduced, the system approaches a double-barrier setup and charge on the sample is closer to quantization, showing charging effects. These charging effects can lead to level reorganization, perhaps the easiest many-body effect to see on scanning tunneling spectroscopy. We have used the theoretical tools to model three classes of example systems, namely graphene ribbons, phthalocyanines, and quantum Hall droplets. On graphene, many-body effects has been shown to be mostly missing, but phthalocyanines showed level reorganization. On the other hand, results on quantum Hall droplets demonstrated the difficulty to capture many-body effects of this strongly correlated system, partially because scanning tunneling spectroscopy at the level treated in this review is a single-particle probe.

Acknowledgements

This research made use of the Aalto Nanomicroscopy Center (Aalto NMC) facilities and was supported by the European Research Council (ERC-2011-StG No. 278698 PRECISE-NANO), the Academy of Finland through its Centres of Excellence Program (projects no. 250280 and 251748), and the Finnish Academy of Science and Letters. We acknowledge the computational resources provided by Aalto Science-IT project and Finland’s IT Center for Science (CSC).

References

[1] G. Binnig, H. Rohrer, Scanning tunneling microscopy, Helv. Phys. Acta 55 (6) (1982) 726–735.
[2] C. J. Chen, Introduction to Scanning Tunneling Microscopy, 2nd Edition, Oxford University Press, Oxford, 2008.
[3] I.-W. Lyo, P. Avouris, Negative differential resistance on the atomic scale: implications for atomic scale devices, Science 245 (4924) (1989) 1369–1371. doi:10.1126/science.245.4924.1369.
[4] B. C. Stipe, M. A. Rezaei, W. Ho, Single-molecule vibrational spectroscopy and microscopy, Science 280 (5370) (1998) 1732–1735. doi:10.1126/science.280.5370.1732.
[5] C. Krull, R. Robles, A. Mugarza, P. Gambardella, Site-and orbital-dependent charge donation and spin manipulation in electron-doped metal phthalocyanines, Nat. Mater. 12 (4) (2013) 337–343. doi:10.1038/nmat3547.
[6] N. Nilius, T. M. Wallis, W. Ho, Development of one-dimensional band structure in artificial gold chains, Science 297 (5586) (2002) 1853–1856. doi:10.1126/science.1075282.
[7] D. Serrate, P. Ferriani, Y. Yoshida, S. W. Hla, M. Menzel, K. von Bergmann, S. Heinze, A. Kubetzka, R. Wiesendanger, Imaging and manipulating the spin direction of individual atoms, Nat. Nano. 5 (5) (2010) 350–353. doi:10.1038/nnano.2010.64.
trolling the charge state of a single redox molecular switch, Phys. Rev. Lett. 106 (2011) 216103. doi:10.1103/PhysRevLett.106.216103

[41] W. Steurer, S. Fatayer, L. Gross, G. Meyer, Probe-based measurement of lateral single-electron transfer between individual molecules, Nat Commun 6, article.

[42] M. Ijás, M. Ervasti, A. Upštū, P. Liljeroth, J. van der Lit, I. Swart, A. Harju, Electronic states in finite graphene nanoribbons: Effect of charging and defects, Phys. Rev. B 88 (2013) 075429. doi:10.1103/PhysRevB.88.075429

[43] D. Toroz, M. Rontani, S. Corni, Visualizing electron correlation by means of ab initio scanning tunneling microscopy images of single molecules, The Journal of Chemical Physics 134 (2011) 024104. doi:10.1063/1.3520567

[44] L. Talirz, P. Ruffieux, R. Fasel, On-surface synthesis of atomically precise graphene nanoribbons, Advanced Materials (2016) n.doi:10.1002/adma.201505738

[45] M. Koch, F. Ample, C. c. Joachim, L. Grill, Voltage-dependent conductance of a single graphene nanoribbon, Nature Nanotechnology 7 (11) (2012) 715–717, cited By 88. doi:10.1038/nnano.2012.169

[46] J. van der Lit, M. P. Boneschanscher, D. Vanmaekelbergh, M. Ijás, A. Upštū, M. Ervasti, A. Harju, P. Liljeroth, I. Swart, Suppression of electron-vibron coupling in graphene nanoribbons contacted via a single atom, Nat. Commun. 4 (2013) 2023. doi:10.1038/ncomms10177

[47] P. Leicht, L. Zielke, S. Bouvron, R. Moroni, E. Voloshina, L. Hammerschmidt, Y. S. Dedkov, M. Fonin, In situ fabrication of quasi-free-standing epitaxial graphene nanoflakes on gold, ACS Nano 8 (4) (2014) 3755–3762. doi:10.1021/nn500396c

[48] A. Kimouche, M. M. Ervasti, R. Drost, S. Halonen, A. Harju, P. M. Joensuu, J. Sainio, P. Liljeroth, Ultra-narrow metallic armchair graphene nanoribbons, Nat. Commun. 6 (2015) 10177. doi:10.1038/ncomms10177

[49] F. Schulz, M. Ijás, R. Drost, S. Hämäläinen, A. Harju, A. Seitsonen, P. Liljeroth, Many-body transitions in a single molecule visualized by scanning tunnelling microscopy, Nature Physics 11 (3) (2015) 229–234, cited By 9. doi:10.1038/nphys3212

[50] F. Schulz, R. Drost, S. K. Hämäläinen, T. Deconinck, A. P. Seitsonen, P. Liljeroth, Epitaxial hexagonal boron nitride on Ir(111): A work function template, Phys. Rev. B 89 (2014) 235429. doi:10.1103/PhysRevB.89.235429

[51] F. Schulz, R. Drost, S. K. Hämäläinen, P. Liljeroth, Templated self-assembly and local doping of molecules on epitaxial hexagonal boron nitride, ACS Nano 7 (12) (2013) 11121–11128. arXiv:http://dx.doi.org/10.1021/nn404840h doi:10.1021/nn404840h

[52] X. H. Qiu, G. V. Nazin, W. Ho, Vibronic states in single molecular electron transport, Phys. Rev. Lett. 92 (2004) 206102. doi:10.1103/PhysRevLett.92.206102

[53] N. S. Wingreen, K. W. Jacobsen, J. W. Wilkins, Inelastic scattering in resonant tunneling, Phys. Rev. B 40 (1989) 11834–11850. doi:10.1103/PhysRevB.40.11834

[54] J. Repp, P. Liljeroth, G. Meyer, Coherent electron-nuclear coupling in oligothiophene molecular wires, Nat Phys 6 (12) (2010) 975–979. doi:10.1038/nphys1802

[55] M. E. Casida, Time-dependent density-functional theory for molecules and molecular solids, J. Mol. Struct.: THEOCHEM 914 (13) (2009) 3–18. doi:10.1016/j.theochem.2009.08.018

[56] Z. F. Ezawa, Quantum Hall Effects, 3rd Edition, WORLD SCIENTIFIC, 2013.

[57] R. B. Laughlin, Anomalous quantum Hall effect: An incompressible quantum fluid with fractionally charged excitations, Phys. Rev. Lett. 50 (1983) 1395–1398. doi:10.1103/PhysRevLett.50.1395

[58] G. Li, A. Luican-Mayer, D. Ablanin, E. Y. Andrei, Evolution of landau levels into edge states in graphene, Nat. Commun. 4 (2013) 1744. doi:10.1038/ncomms2767