Dynamical image-charge effect in molecular tunnel junctions
Beyond energy level alignment

Jin, Chengjun; Thygesen, Kristian Sommer

Published in:
Physical Review B. Condensed Matter and Materials Physics

Link to article, DOI:
10.1103/PhysRevB.89.041102

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Dynamical image-charge effect in molecular tunnel junctions: Beyond energy level alignment

Chengjun Jin and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark

(Received 11 October 2013; revised manuscript received 18 December 2013; published 6 January 2014)

When an electron tunnels between two metal contacts it temporarily induces an image charge (IC) in the electrodes which acts back on the tunneling electron. It is usually assumed that the IC forms instantaneously such that a static model for the image potential applies. Here we investigate how the finite IC formation time affects charge transport through a molecule suspended between two electrodes. For a single-level model, an analytical treatment shows that the conductance is suppressed by a factor $Z^2$, where $Z$ is the quasiparticle renormalization factor, compared to the static IC approximation. We show that $Z$ can be expressed either in terms of the plasma frequency of the electrode or as the overlap between electrode wave functions corresponding to an empty and filled level, respectively. First-principles GW calculations for benzene-diamine connected to gold electrodes show that the dynamical corrections can reduce the conductance by more than a factor of two when compared to static GW or density functional theory where the molecular energy levels have been shifted to match the exact quasiparticle levels.

DOI: 10.1103/PhysRevB.89.041102 PACS number(s): 85.65.+h, 31.70.Dk, 71.10.+w, 73.20.–r

The effect of image forces on tunneling electrons was first studied by Sommerfeld and Bethe [1] and Holm [2] in the 1930s, and later refined by Simmons [3] to a form, which still today is widely used. In Simmons model, the effect of image forces is described by a simple $1/z$ correction to the tunneling barrier. Its range of validity has recently been critically examined on basis of ab initio calculations and experimental data for (sub-)nanometer-sized tunneling junctions [4–8].

Image charge (IC) forces also have important consequences for electron transport at metal-molecule interfaces because they influence the position of the molecular energy levels relative to the metal Fermi level [9–14]. Because the interaction with the image charge lowers the energy cost of adding an electron/hole to a molecular orbital, the occupied energy levels are shifted upwards, while the empty levels are shifted downwards in energy as the molecule approaches a metal surface.

Theoretically, the image forces are challenging to describe because they are created by the electron on which they act. To properly include such correlation effects one must go beyond standard single-particle theories like Hartree-Fock and density functional theory (DFT) [15,16]. For transport in molecular junctions, this has been done previously using the GW approximation to the electron self-energy both in the steady state [17–20] and time-dependent [21] regimes. Due to the computational complexity of such many-body methods, simple ad hoc correction schemes have been developed which shift the energy of the molecular orbitals by an amount estimated from a classical image charge model [22,23]. Such correction schemes, generally termed DFT + $\Sigma$, have been shown to improve the agreement with experiments compared to the uncorrected DFT result [24]. An interesting question is then whether such a level correction scheme captures all the effects of the IC on electron transport if the corrections are chosen to reproduce the exact level alignment for the frontier orbitals. It was recently shown that the IC not only influences the energy of the molecular orbitals but also their spatial shape [25]. A change in orbital shape will change the hybridization with the metal states and thereby affect the tunneling rate.

This effect is beyond the DFT + $\Sigma$ schemes, but should be significant only for highly polarizable molecules.

Except for the few many-body calculations, all previous attempts to model the IC effect in molecular transport junctions have been based on the assumption that the IC forms instantaneously such that a static IC model applies. On the other hand, it is intuitively clear that the role of the IC depends on the time it takes to polarize the electrode compared to the time the electron spends on the molecule. The former is given roughly by the inverse plasmon frequency of the electrode, $\tau_p \approx 1/\omega_p$, while a simple expression for the latter follows from the time-energy uncertainty relation, $\tau_{\text{run}} \approx \hbar/|E_F - \epsilon_0|$, where $\epsilon_0$ is the energy of the molecular orbital closest to the Fermi level. We note that the related problem of how a finite plasmon frequency influences the spatial form of the image potential at a metal surface has been studied by several authors in the past [26–28].

In this Rapid Communication we show, using both a simple one-level model and first-principles many-body calculations, that the finite electrode response time always suppresses the conductance of a molecular junction compared to the result of a noninteracting model with the exact same level alignment (static IC approximation). Formally this is a consequence of the reduction of the quasiparticle weight of the molecular resonance from 1 to $Z < 1$ due to the electron-electron interactions which shift spectral weight from the single-particle excitation to other excitations (in particular plasmons). In the off-resonance tunneling regime, the conductance of the one-level model is suppressed by $Z^2$ compared to the static result. We provide two complementary physical explanations for this reduction. In a dynamical picture, it can be related to the ratio between the characteristic IC formation time $\tau_p$ and the dwell time of the electron on the molecule expressing the reduced screening of the electron due to the “lagging behind” of the IC. In a picture of hopping between many-body states, $Z$ can be expressed as an overlap of the electrode wave function with and without the IC and thus explains the origin of the reduced tunneling rate as a mismatch between the initial and final states of the electrode. Ab initio GW calculations for benzene-diamine (BDA) connected to gold electrodes shows...
a conductance reduction of almost a factor 3 compared to the static approximation (noninteracting transport through optimally tuned energy levels), demonstrating the importance of dynamical corrections for realistic systems.

We consider the problem of electron transport through a single electronic level \(|a\rangle\) coupled to left (L) and right (R) electrodes. Due to the hopping matrix elements between \(|a\rangle\) and the states of the electrodes, the level is broadened into a resonance with a finite spectral width \(\gamma\), which we take to be energy independent for simplicity. We assume that the level is unoccupied, i.e., \(\epsilon_a > E_F + \gamma\), however, the case of an occupied level is treated completely analogously. The time-ordered Green’s function of the localized level can be written

\[
G_A(\omega) = \frac{1}{[\omega - \epsilon_a - \text{Re} \Sigma_A(\omega)] + i[\gamma - \text{Im} \Sigma_A(\omega)]},
\]

where the self-energy \(\Sigma_A(\omega) = \langle |a \rangle \Sigma(\mathbf{r}, \mathbf{r}’, \omega) |a\rangle\) accounts for the Coulomb interaction between electrons in the electrodes and an electron in \(|a\rangle\). To lowest order in the interaction, the self-energy contains the Hartree and exchange potentials of Hartree-Fock theory. These terms do not contribute to the image charge effect and are therefore absorbed in \(\epsilon_a\). Thus \(\Sigma_A\) includes only the higher order terms (correlation effects).

The screening from the electrodes shifts the pole of the GF from \(\epsilon_a\) to the quasiparticle (QP) energy

\[
\epsilon_a^\text{QP} = \epsilon_a + \Delta \epsilon_c, \quad \Delta \epsilon_c = Z \Sigma_A(\epsilon_a),
\]

where \(\Delta \epsilon_c\) denotes the image charge shift and \(Z = [1 - d \Sigma_A(\epsilon_a)/d\omega]^{-1}\) is the renormalization factor to be discussed later.

Within the GW approximation [29], the self-energy takes the form

\[
\Sigma(\mathbf{r}, \mathbf{r}’, \omega) = \frac{i}{2\pi} \int G_0(\mathbf{r}, \mathbf{r}’, \omega + \omega’) \tilde{W}(\mathbf{r}, \mathbf{r}’, \omega’) d\omega’,
\]

where \(\tilde{W} = W - \nu\) and \(W\) is the dynamically screened Coulomb interaction. We have subtracted the bare Coulomb interaction \(\nu = 1/|\mathbf{r} - \mathbf{r}’|\) from \(W\) to avoid double counting of the exchange energy which is already contained in \(\epsilon_a\). The unperturbed Green’s function, is given by

\[
G_0(\mathbf{r}, \mathbf{r}’, \omega) = \frac{\psi_a(\mathbf{r})\psi_a^*(\mathbf{r}’)}{\omega - \epsilon_a + i\gamma} + \sum_k \frac{\psi_k(\mathbf{r})\psi_k^*(\mathbf{r}’)}{\omega - \epsilon_k + i0^+ \text{sgn}(\epsilon_k - E_F)}.
\]

In terms of the density response function of the metal electrode \(\chi\), we have (suppressing the integration over spatial variables) \(\tilde{W}(\omega) = \nu \chi(\omega)\nu\). Neglecting the spatial overlap between \(|a\rangle\) and the metal states, the relevant matrix element of the screened interaction \(\langle |a\rangle \tilde{W}(\omega) |a\rangle\) becomes

\[
\tilde{W}_a(\omega) = \int \int V_a(\mathbf{r}’) \chi(\mathbf{r}, \mathbf{r}’, \omega) V_a(\mathbf{r}) d\mathbf{r} d\mathbf{r}’,
\]

where \(V_a(\mathbf{r})\) is the potential created by an electron in the state \(|a\rangle\),

\[
V_a(\mathbf{r}) = \int \frac{|\psi_a(\mathbf{r}’)|^2}{|\mathbf{r} - \mathbf{r}’|} d\mathbf{r}’.
\]

A Feynman diagram of the self-energy is shown in Fig. 1.

FIG. 1. (Color online) Upper panel: When the localized level \(|a\rangle\) is empty, the charge distribution corresponding to the electrode ground state, \(|\Psi_0\rangle\), is homogeneous (no image charge). Lower panel: When the level is occupied, the potential from the localized electron, \(V_a(\mathbf{r})\), induces an image charge in the electrode ground state, \(|\Psi_0\rangle\). A Feynman diagram for the self-energy describing the IC effect is shown in the upper panel.

Using a plasmon pole approximation (PPA) for the response function,

\[
\tilde{W}_a(\omega) = A \left( \frac{1}{\omega - \omega_p + i\gamma_p} - \frac{1}{\omega + \omega_p - i\gamma_p} \right),
\]

the self-energy can be evaluated using complex contour integration

\[
\Sigma_a(\omega) = A \left( \frac{1}{\omega - \epsilon_a - \omega_p + i(\gamma + \gamma_p)} \right).
\]

where \(\omega_p\) and \(\gamma_p\) are the characteristic plasmon energy and spectral width, respectively. It follows that the imaginary part \(\Delta \epsilon_c = Z \Sigma_a(\epsilon_a)\) in the range between \(E_F\) and \(\epsilon_a\), this means we can set \(\Gamma = 0\) in Eq. (8). The constant \(A\) can be fixed by invoking the condition \(\Delta \epsilon_c = Z \Sigma_a(\epsilon_a)\), which results in

\[
A = \frac{\Delta \epsilon_c \omega_p}{\omega_p - \Delta \epsilon_c}.
\]

Close to equilibrium, i.e., for small bias voltages, the conductance is given by Landauer’s formula \(G = \frac{2e^2}{h} T(E_F)\) [30]. For the single-level model, the transmission at the Fermi level can be written

\[
T(E_F) = \frac{\gamma^2}{(E_F - \epsilon_a^{\text{eff}})^2 + \gamma^2},
\]

where we have defined the effective energy level seen by the tunneling electron as

\[
\epsilon_a^{\text{eff}} = \epsilon_a + \text{Re} \Sigma_a(E_F) = \epsilon_a + \Delta \epsilon_c \left( \frac{\omega_p}{|E_F - \epsilon_a| + \omega_p} \right).
\]
In the above expression we have assumed, for simplicity of the expression, that $\Delta \varepsilon_{\text{ic}} \ll \omega_p$. The transmission through the interacting level is thus equivalent to transmission through a noninteracting level with energy $\varepsilon_a^{\text{QF}}$. When the image charge formation is fast compared to the average time spent by the electron on the molecule, i.e., when $\omega_p \gg |E_F - \varepsilon_a|$, the effective level equals $\varepsilon_a^{\text{QF}}$ and the static image charge approximation is valid. In the opposite regime where the tunneling time is short compared with the image charge formation is fast compared to the average time spent by the electrode response time into an effective level position.

Within the quasiparticle approximation, one expands the electrode, i.e., when $\Delta \varepsilon_{\text{ic}} \ll \omega_p$, one obtains

$$Z = 1 - \frac{\Delta \varepsilon_{\text{ic}}}{\omega_p}. \quad (12)$$

Noting that the normalized final state is $|\tilde{\Psi}_0\rangle = (|\Psi_0\rangle + |\Psi_0^{(1)}\rangle)/(1 + (|\Psi_0^{(1)}\rangle|^2)^{1/2}$, and comparing with Eq. (12), it follows that

$$Z = (|\tilde{\Psi}_0\rangle|\Psi_0\rangle)^2. \quad (16)$$

In fact, this also follows from a more general result stating that $Z$ is the squared norm of the QP state $|\alpha\rangle$ (see, e.g., Ref. [25]). Equation (16) shows that the origin of the $Z^2$ conductance suppression expressed by Eq. (13) (at least in the cotunneling regime where $|E_F - \varepsilon_a^{\text{QF}}| \gg \gamma$), can be understood as a mismatch of the initial and final states of the electrodes. Here we note the similarity with the phenomenon known as Franck-Condon blockade where transport through a molecule is suppressed due to reduced overlap between the initial and final vibronic states of the molecule [31]. According to Eq. (16) the magnitude of $Z$ is determined by the relative weight of the component $|\Psi_0^{(1)}\rangle$ in the final state $|\tilde{\Psi}_0\rangle$. We can relate the norm of $|\Psi_0^{(1)}\rangle$ to the response time of the electrode by noting that the terms in Eq. (14) have an $E_s - E_0 \approx \omega_p$. Thus a faster electrode response, i.e., larger $\omega_p$, is equivalent to a smaller perturbation of the ground state and thus $Z$ closer to unity. This is again consistent with Eq. (12).

To test the role of dynamical screening under more realistic conditions, we have performed first-principles GW calculations for the benchmark system of BDA connected to gold electrodes (see inset of Fig. 2). The details of the calculation follow Ref. [18]. In brief, the Green’s function of the contacted molecule is obtained by solving the Dyson equation self-consistently including both lead coupling self-energies and the GW self-energy. We use a basis of numerical atomic orbitals
at the double-$\zeta$ plus polarization level for the gold electrodes and double-$\zeta$ for the BDA. The GW self-energy is evaluated in a spatial region containing the molecule and the four closest Au atoms on each side of the molecule. For the considered junction geometry this is sufficient because the IC is essentially confined to the tip Au atoms [18].

In Fig. 2 we show the transmission function calculated using four different methods. In addition to the GW result we show the transmission obtained from DFT with the standard Perdew-Burke-Ernzerhof exchange-correlation (xc) functional [32]. Not surprisingly the latter yields a higher conductance due to the well-known underestimation of the molecular energy gap. To isolate the role of dynamical effects we have used a “scissors operator” to adjust the energies of the molecular orbitals in the DFT calculation to those obtained with GW:

$$\Sigma_{SO} = \sum_{\nu \mu} \Delta \varepsilon_{\nu} |\psi_{\nu}\rangle \langle \psi_{\nu}|.$$

(17)

The molecular orbitals $|\psi_{\nu}\rangle$ are obtained by diagonalizing the DFT Hamiltonian within the subspace spanned by the basis functions of the BDA. In practice, the energy shift ($\Delta \varepsilon_{\nu}$) of the three highest occupied and three lowest unoccupied orbitals are fitted to match the positions of the main peaks in the GW transmission spectrum. As a fourth method we followed the QPscGW scheme of Schilfgaarde et al. to construct a static and Hermitian xc potential from the GW self-energy using the expression [33]

$$V^{xc} = \frac{1}{2} \sum_{\nu \mu \in \text{mol}} |\psi_{\nu}\rangle \text{Re} \left[ \left( \Sigma (\varepsilon_{\nu}^{QP}) \right)_{\nu \mu} + \left[ \Sigma (\varepsilon_{\nu}^{QP}) \right]_{\nu \mu} \right] |\psi_{\mu}\rangle,$$

(18)

with the QP energies $\varepsilon_{\nu}^{QP}$ obtained from the full GW calculation. As can be seen from Fig. 2, the QPscGW and DFT + $\Sigma_{SO}$ methods yield very similar transmission spectra. This is because the off-diagonal matrix elements of $V^{xc}$ from Eq. (18) are essentially zero, meaning that the DFT and QP molecular orbitals coincide. (This is not surprising given the low polarizability of BDA [25].) We thus conclude that the observed difference in transmission between full GW on the one hand and DFT + $\Sigma_{SO}$ or QPscGW on the other hand, is neither due to differences in energy level alignment nor in the spatial shape of orbitals, but originates from the frequency dependence of the GW self-energy.

In Fig. 3 we show the Hartree-Fock (HF) and GW results for the spectral function of the BDA highest occupied molecular orbital (HOMO) together with the imaginary and real parts of the GW self-energy $|\psi_{H}| \Sigma (\omega) |\psi_{H}|$. From this it follows that the GW self-energy shifts the HOMO up in energy by 1.9 eV. The corresponding self-energy shift for BDA in the gas phase, caused by intramolecular screening, is found to be 1.0 eV. From this we conclude that the size of the IC shift, caused by the metallic screening, is 0.9 eV. It is clear from the almost linear behavior of $\text{Re} \Sigma (\omega)$, that the linear expansion of $\Sigma$ leading to Eq. (13) is well justified. Furthermore, the imaginary part of the GW self-energy vanishes for energies above $\varepsilon_{H}$ in agreement with the one-level model. The width of the spectral functions in Fig. 3 is given by the imaginary part of the coupling self-energy (not shown). The energy variation of this broadening follows the density of states at the gold tip atom. This explains the larger broadening of the GW resonance compared to the HF resonance which is situated below the gold $d$ band.

From the slope of $\text{Re} \Sigma$ we obtain the renormalization factor of $Z = 0.84$. Based on the one-level model this should lead to a conductance suppression by a factor $Z^{2} = 0.71$ which is, however, not sufficient to explain the observed difference between the GW and static GW result (see Fig. 2). The reason for this is that the BDA junction is not well described by a one-level model. While the unoccupied states play a minor role for the conductance, the HOMO-2, which is an antibonding version of the HOMO, must be included to obtain a realistic model. This points to a nontrivial interplay between the dynamical effects on different transport channels.

In conclusion, our results demonstrate that the role of electron-electron interactions in charge transport across a metal-molecule interface goes beyond the well-established effect on the energy level alignment. In general, the image charge dynamics renormalizes the level broadening (or equivalently the tunneling rate) by an amount that depends on the plasmon frequency of the electrode. Since the former can be tuned, e.g., by nanostructuring or electrostatic gating, this could provide a basis for experimental investigations of the dynamical image charge effect.

The authors acknowledge support from the Danish Council for Independent Research’s Sapere Aude Program through Grant No. 11-1051390.


