

# Hot electron cascades in the scanning tunneling microscope

Natalia L. Schneider,<sup>1</sup> Peter Johansson,<sup>2</sup> and Richard Berndt<sup>1</sup><sup>1</sup>*Institut für Experimentelle und Angewandte Physik, Christian-Albrechts-Universität zu Kiel, 24098 Kiel, Germany*<sup>2</sup>*School of Science and Technology, Örebro University, 70182 Örebro, Sweden and Department of Applied Physics, Chalmers University of Technology, 41296 Göteborg, Sweden*

(Received 21 November 2012; published 9 January 2013)

The nonequilibrium distribution of electrons at the junction of a scanning tunneling microscope is investigated by detecting photons with energies  $h\nu > eV$ , where  $V$  is the bias voltage. Electrons are found at energies exceeding the Fermi level by almost  $eV$ . While their distribution deviates from a Fermi-Dirac function it is consistent with a model of hot electrons and holes that diffuse in energy and real space.

DOI: [10.1103/PhysRevB.87.045409](https://doi.org/10.1103/PhysRevB.87.045409)

PACS number(s): 78.68.+m, 68.37.Ef, 73.20.Mf, 72.10.-d

## I. INTRODUCTION

The tunneling current in a scanning tunneling microscope (STM) locally deposits energy in the electron systems of the tip and the sample. Hot electrons are injected into the positive electrode leaving behind hot holes in the negative electrode. The hot carriers may rapidly distribute their energy in the electron system via electron-electron interaction. Moreover, phonons may be created and transport the energy away from the tunneling contact. Usually the heating of the electron systems is insignificant and the STM provides information on the equilibrium density of states. However, this approximation may break down when elevated currents are used to probe the quantum transport through atomic or molecular contacts.<sup>1-6</sup>

The nonequilibrium distribution of electrons and relaxation rates have been probed with an electrode coupled by tunneling to biased metallic wires<sup>7</sup> or carbon nanotubes<sup>8</sup> in a three-terminal arrangement. In a conventional two-terminal STM thermovoltages may be observed.<sup>9</sup> Accessing the energy distribution of hot carriers, however, is difficult because the tunneling current represents an integral over all electronic states. In contrast, photons emitted from the STM carry more direct information on excited states. In the absence of electron-electron interaction, at low temperatures, the highest photon energy is limited by the voltage between the electrodes,  $h\nu < eV$ . Photon energies exceeding  $eV$  indicate that hot carriers are involved.<sup>10-12</sup> For simplicity, we refer to this emission as  $2e$  light.

Here, we investigate  $2e$  light that is emitted from a Au(111) surface. Spectra of this light vary with bias and are shown to reflect the nonequilibrium distribution of charge carriers at the junction. The distributions determined from the experiments, which deviate from thermal distributions, in particular at the high-energy end, are analyzed using model calculations. We find hot carrier occupation numbers of the order of  $10^{-4}$  (corresponding to carrier concentrations of the order  $10^{18} \text{ cm}^{-3} \text{ eV}^{-1}$ ) at a junction current of  $10 \mu\text{A}$ .

On metal surfaces, the high-frequency noise of the tunneling current may excite localized plasmons, which in turn emit light.<sup>14,15</sup> Emission at  $h\nu$  slightly exceeding  $eV$  was first reported from Au-Au junctions at ambient temperature and fairly low currents ( $10 \text{ nA}$ ).<sup>16</sup> While an elevated temperature of the electron gas ( $1200 \text{ K}$ ) was suggested to cause smearing of the Fermi distribution, no detailed analysis of the electron distribution was presented. The emission from

Au-Au junctions in the contact regime was interpreted in terms of blackbody radiation of electron gas.<sup>17</sup> However, the electromagnetic properties of the junction, namely the presence of tip-induced plasmon (TIP) modes,<sup>18</sup> which affect the spectral characteristics, were not taken into account. While the present work focuses on a metallic model system it should be noted that fluorescence at  $h\nu > eV$  has also been reported from molecular films.<sup>19-21</sup>

The thermalization of electrons in metals has been investigated with pulsed-laser excitation. Electron-electron interaction leads to thermalization of the electron system on a sub-ps timescale over which the distribution evolves from a nonthermal to a hot Fermi-Dirac shape.<sup>22</sup> However, thermalization can be slowed down in the presence of energy transfer to the lattice.<sup>22,23</sup> A nonthermalized electron distribution was reported for a low density of excited electrons ( $< 2 \times 10^{18} \text{ cm}^{-3}$ ).<sup>24</sup>

## II. EXPERIMENT

### A. Details

The experiments were performed with a STM operating in ultrahigh vacuum at low temperature ( $5.8 \text{ K}$ ). Light emitted from STM was collected with a lens and guided to a grating spectrometer and a liquid nitrogen cooled CCD camera via an optical fiber.<sup>25</sup> Spectra are not corrected for the efficiency of optical setup, which is shown in Fig. 1(d). Au(111) surfaces and W tips were cleaned in UHV by  $\text{Ar}^+$  bombardment and annealing cycles. Finally the tips were indented into the Au(111) substrate to increase their stability for contact experiments.

### B. Results

Figure 1 shows luminescence spectra acquired in series with the same tip above a Au(111) surface. Figure 1(a) was measured in the tunneling range at low current. An elevated bias voltage  $V = 3.5 \text{ V}$  was used to ensure that the shape of the spectrum essentially reflects the shape of the tip-induced plasmon mode. At low photon energies, however, the spectral shape is affected by vanishing response of the detection system, which is indicated by a dashed line in Fig. 1(d). As the tunneling current is low, heating of the electron gas is insignificant. The two data sets shown in blue and red colors were recorded before and after the measurements of Figs. 1(b)–1(d). Their close similarity confirms that no

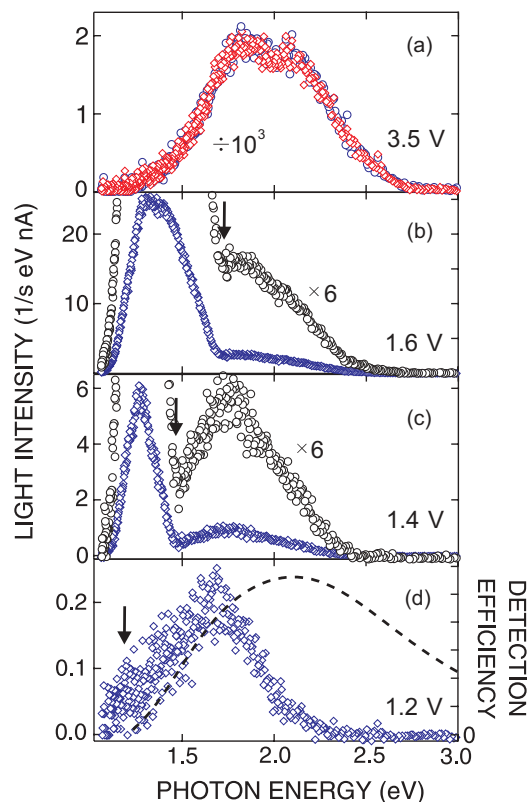


FIG. 1. (Color online) (a) Luminescence spectra recorded with a Au coated W tip on Au(111) in the tunneling regime at  $V = 3.5$  V and  $I = 1$  nA before (blue) and after (red) the contact data of (b)–(d) was measured. Black arrows indicate photon energies  $h\nu = eV$ . (b)–(d) Spectra acquired at sample voltages  $V = 1.6, 1.4$ , and  $1.2$  V and an elevated current  $I = 10$   $\mu$ A. A dashed curve in (d) shows an estimate of the detection efficiency of the optical setup.

significant changes of the tip occurred during the experiments, which would affect the spectra.<sup>26–28</sup>

The spectra of Figs. 1(b)–1(d) were recorded at  $I = 10$   $\mu$ A and sample voltages  $V = 1.6, 1.4$ , and  $1.2$  V. The corresponding conductances are close to that of a single atom contact on Au.<sup>11,29</sup> At the low bias voltages used, the limited phase space for inelastic transitions limits the intensities at high photon energies. The threshold for single-electron processes ( $1e$  light),  $h\nu = eV$ , is indicated by arrows. The spectra acquired at  $1.6$  and  $1.4$  V exhibit intense  $1e$  light while the related signal is weak at  $V = 1.2$  V owing to low detection efficiency [cf. dashed line in Fig. 1(d)]. In agreement with previous work,<sup>11</sup> additional significant emission occurs at higher photon energies,  $h\nu > eV$ , which involves hot charge carriers.

While the spectra of Fig. 1 reflect the presence of hot electrons or holes they are also influenced by other factors such as the electromagnetic modes of the tip–sample junction. Following Ref. 30 we use the ratio  $N_2(h\nu)/N_1(h\nu)$ , where  $N_2$  and  $N_1$  are the intensities from spectra recorded at low ( $V_2 = 1.2, 1.4$ , and  $1.6$  V) and high ( $V_1 = 3.5$  V) bias and different currents, respectively, to obtain a quantity that is more closely related to calculated electron distributions. In particular, the variation of the detection efficiency of the optical setup with the wavelength cancels out. Figure 2 shows light emission spectra for photons with  $h\nu > eV_2$  at  $V_2 = 1.2, 1.4$ , and  $1.6$  V divided

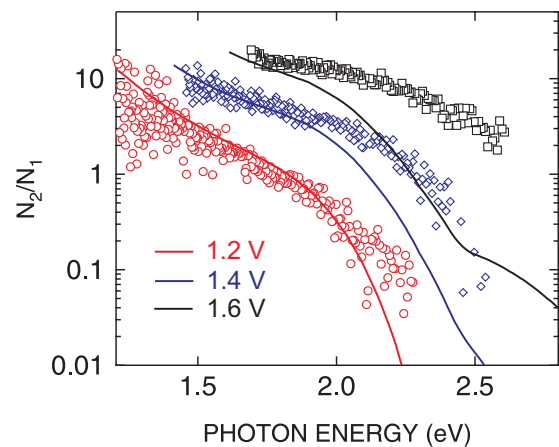


FIG. 2. (Color online)  $2e$ -intensity  $N_2$  at sample voltages  $V_2 = 1.2, 1.4$ , and  $1.6$  V and a tunneling current  $I_2 = 10$   $\mu$ A (red circles, blue diamonds, and black rectangles, respectively) divided by the  $1e$ -intensity  $N_1$  at  $V_1 = 3.5$  V and  $I_1 = 1$  nA. Photon energies  $h\nu > eV_2$  are shown. Calculated values of the ratio  $N_2/N_1$ , multiplied by a factor of 3, are shown with solid lines.

by the spectrum at  $V_1 = 3.5$  V. As expected for the density of hot electrons,  $N_2/N_1$  increases at higher bias voltages.

### III. CALCULATIONS

#### A. Model

We have carried out model calculations of the intensity of  $1e$  and  $2e$  light in order to get a comparison with the experimental results. In general, the light emission intensity from the STM junction, by applying the Fermi golden rule, can be written as a sum over initial and final states

$$\frac{d^2 P}{d\Omega d(h\nu)} = \frac{v^2 |G(h\nu)|^2}{2\varepsilon_0 c^3} \sum_{i,f} |j_{if}|^2 f_e(E_i) \times [1 - f_c(E_f)] \delta(E_i - E_f - h\nu). \quad (1)$$

Here  $G(h\nu)$  denotes the enhancement of the electromagnetic field in the tunneling gap between sample and tip at photon energy  $h\nu$ , and  $j_{if}$  is a transition matrix element for the inelastic tunneling transition from an initial electron state of energy  $E_i$  in the emitter electrode to a final state of energy  $E_f$  in the collector. The distribution functions  $f_e$  and  $1 - f_c$  appearing in the sum dictate that an inelastically tunneling electron must start from an occupied state in the emitter and make a transition to an unoccupied one in the collector. For the case of  $1e$ -light emission, it is sufficient to view the electrodes as two Fermi seas at temperature  $T = 0$  where the emitter Fermi energy lies  $eV$  above the collector Fermi energy. The  $2e$ -light emission, on the other hand, is due either to hot electrons above the emitter Fermi level,  $f_e > 0$ , tunneling into unoccupied states in the collector or hot holes below the collector Fermi level,  $1 - f_c > 0$ , being filled by electrons tunneling from the emitter. We wish to stress here that the distribution functions  $f_e$  and  $f_c$ , as the experimental as well as theoretical results will show, are not identical to the Fermi distribution function.

Our calculation of the  $1e$ -light emission follows earlier work.<sup>14,18</sup> The tunneling current and inelastic tunneling matrix elements are calculated employing an essentially

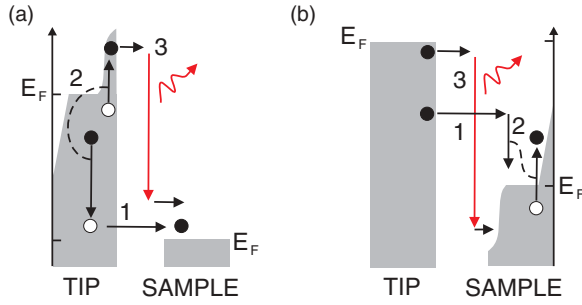


FIG. 3. (Color online) Mechanisms leading to the emission of  $2e$  light. Gray shading indicates the occupation electron states. (a) Elastic tunneling (1) at energies below the Fermi level of the emitter creates a hot hole. This causes a cascade (2) and results in electrons with energies above the Fermi level of the tip. Inelastic tunneling of these electrons leads to the emission of photons with energies  $h\nu > eV$ . (b) A hot electron is injected into the sample (1). A cascade (2) produces hot holes. Inelastic tunneling to this hole results in  $2e$  light (3). The modeling presented in the text predicts similar probabilities for both processes.

one-dimensional model, whereas the electromagnetic field enhancement in the cavity between tip and sample is evaluated for a spherical model tip ( $R = 30$  nm) with the dielectric properties of tungsten.

To deal with the  $2e$ -light emission we calculate the distribution functions appearing in Eq. (1) within a simple model of hot-hole-hot-electron cascades (Fig. 3). The rate of the energy transfer processes in the model from hot carriers to electrons in the Fermi sea are set by phase-space considerations.<sup>31</sup> At the same time all of the hot holes and electrons are allowed to diffuse in the electrodes. Thus, the model leads to a combined diffusion and energy transfer rate problem which we solve in order to determine the quantity of primary interest for the photon emission at energies  $h\nu > eV$ , namely the hot electron occupation  $f_e$  above the tip Fermi level and the hot hole occupation  $1 - f_c$  below the sample Fermi level. The model as well as the parameter values employed are discussed in more detail in the Appendix.

## B. Results

The results of such a “cascade” calculation are shown in Fig. 4. The distributions are to a large extent each other’s mirror images around the Fermi level. The main difference occurs for the injected excitations. Since the tunneling probability is the largest for electrons with high energy, it is easier to inject hot electrons far from the Fermi level in the sample than to inject hot holes far from the Fermi level in the tip. The quantities relevant to the  $2e$ -light emission  $f_e$  and  $1 - f_c$  here reach values in the range of  $10^{-5}$ – $10^{-4}$  for energies some 0.5–1 eV away from the Fermi level.

## IV. DISCUSSION

To compare the calculated occupation numbers with experimentally accessible quantities we note that the ratio  $N_2/N_1$

$$\frac{N_2(V_2, I_2)}{N_1(V_1, I_1)} = \frac{[d^2 P/d(h\nu)d\Omega](V_2, I_2)}{[d^2 P/d(h\nu)d\Omega](V_1, I_1)} \quad (2)$$

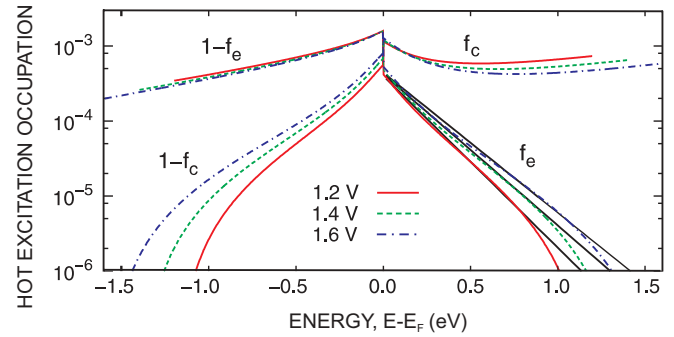


FIG. 4. (Color online) Calculated hot electron ( $f$ ) and hot hole ( $1 - f$ ) occupation probabilities at a tunneling contact at three bias voltages, and for a tunneling current  $I = 10 \mu\text{A}$ . Data is shown for the emitter (index “e”) and collector (index “c”) electrodes. Straight lines show the high-energy tails of distributions  $\propto \exp[(E_F - E)/(k_B T)]$  at temperatures 2200, 2500, and 2700 K. A local energy scale for each electrode is used.

between  $2e$  and  $1e$ -light emission at the same photon energy but with different bias voltages and tunnel currents obtained from Eq. (1) is primarily determined either by  $f_e$  or  $1 - f_c$ . Results for  $N_2/N_1$  are shown in Fig. 2 for  $V_1 = 3.5$  V,  $I_1 = 1$  nA,  $I_2 = 10 \mu\text{A}$ , and  $V_2$  set to 1.2, 1.4, and 1.6 V, respectively, together with experimental results for the same ratio. The theoretical and experimental results show the same dependence on the photon energy, with an initially steady slope (on a logarithmic scale) followed by a much faster decrease closer to the maximum photon energy that can possibly be obtained from the hot excitations. This behavior of  $N_2/N_1$  is different from that of a hot electron Fermi gas, where the rapid decrease at the high-energy end of the spectrum would not be present. Neglecting their high energy part, the distributions of hot electrons correspond to temperatures of 2200–2700 K (Fig. 4). The ratio between the two intensities reaches numerical values of order 1. This agrees well with what one can expect from the calculated hot-excitations occupation numbers. While the occupation numbers are of the order  $10^{-4}$ , with a current that is a factor of  $10^4$  higher for the  $2e$ -light emission the probability of an electron crossing the tunnel barrier is roughly  $10^4$  times larger in both elastic and inelastic tunneling events, hence the similar intensities in the two cases.

A closer look at the results for  $N_2/N_1$  in Fig. 2 shows a good agreement between theory and experiment for  $V = 1.2$  V, while there are larger differences for 1.4 and 1.6 V. There the hot electron and hole occupation factors are not the only factor determining  $N_2/N_1$  within the model. Instead, in the calculated results the electromagnetic response  $|G|^2$  still adds some more features to  $N_2/N_1$  that cannot be seen in the hot-excitations distributions alone. Since the  $2e$ -light emission occurs at elevated tunneling currents, which correspond to a smaller gap between the tip and sample, yielding a larger resonant electromagnetic enhancement than at  $I = 1$  nA. At higher photon energies, above about 2.2 eV, the electromagnetic enhancement found in the model is less geometry sensitive, which causes a drop in the calculated  $N_2/N_1$  ratio with increasing  $h\nu$  in addition to what the decrease in hot carrier occupation gives. The experimental results (for 1.4 and 1.6 V) do not exhibit this decrease, and in fact show much the

same dependence on energy as the calculated hot carrier distributions do. We do not fully understand the reason for these differences between experiment and theory. A possible explanation is that the experimental tip shape is sufficiently different from the model tip that geometry effects on the electromagnetic response are weaker in the experiment. Model calculations neglecting the calculated shift of the plasmon resonance actually lead to a slightly better agreement with the experimental data.

## V. CONCLUSIONS

In conclusion, we used spectra of the light that is emitted from a biased STM junction to determine the distribution of hot electrons at the tip apex. Electrons are found at energies exceeding the Fermi level by up to  $eV$ . While their distribution deviates from a Fermi-Dirac function it is consistent with a model of hot electrons and holes which diffuse in energy and real space.

## ACKNOWLEDGMENT

This work was funded by the Deutsche Forschungsgemeinschaft through project GR 1409/5-1.

## APPENDIX: MODEL FOR HOT-ELECTRON-HOLE CASCADE

To calculate the concentration of hot electrons and holes at the STM tunneling contact that can cause light emission at photon energies  $h\nu > eV$ , we employ a model that takes into account hot-electron-hole cascades, i.e., essentially diffusion in energy space, and diffusion in real space on an equal footing. The diffusion in energy space is dealt with along the lines of the work of Ritchie.<sup>31</sup> We denote the concentration of hot electrons with an energy of  $\varepsilon > E_F$  at the point  $\vec{r}$  with  $u(\vec{r}, \varepsilon)$  and the concentration of hot holes with energy  $\varepsilon < E_F$  with  $v(\vec{r}, \varepsilon)$ . These quantities are governed by diffusion equations which in a steady-state situation ( $\frac{\partial u}{\partial t} = 0$  and  $\frac{\partial v}{\partial t} = 0$ ) take the form

$$\begin{aligned} u/\tau_0 - D\nabla^2 u &= S_e(\vec{r}, \varepsilon) - u(\varepsilon) \int_{E_F}^{\varepsilon} \mu_{e1}(\varepsilon, \varepsilon') d\varepsilon' \\ &+ \int_{\varepsilon}^{E_{\max}} u(\varepsilon') [\mu_{e1}(\varepsilon', \varepsilon) + \mu_{e2}(\varepsilon', \varepsilon)] d\varepsilon' \\ &+ \int_{E_{\min}}^{2E_F - \varepsilon} v(\varepsilon') \mu_{eh}(\varepsilon', \varepsilon) d\varepsilon', \end{aligned} \quad (\text{A1})$$

and

$$\begin{aligned} v/\tau_0 - D\nabla^2 v &= S_h(\vec{r}, \varepsilon) - v(\varepsilon) \int_{\varepsilon}^{E_F} \mu_{h1}(\varepsilon, \varepsilon') d\varepsilon' \\ &+ \int_{E_{\min}}^{\varepsilon} v(\varepsilon') [\mu_{h1}(\varepsilon', \varepsilon) d\varepsilon' + \mu_{h2}(\varepsilon', \varepsilon)] d\varepsilon' \\ &+ \int_{2E_F - \varepsilon}^{E_{\max}} u(\varepsilon') \mu_{eh}(\varepsilon', \varepsilon) d\varepsilon'. \end{aligned} \quad (\text{A2})$$

Here  $E_{\min}$  is the lowest possible hole energy and  $E_{\max}$  the highest possible electron energy given the energy distribution of the injected electrons and holes. The source terms  $S_e$  and

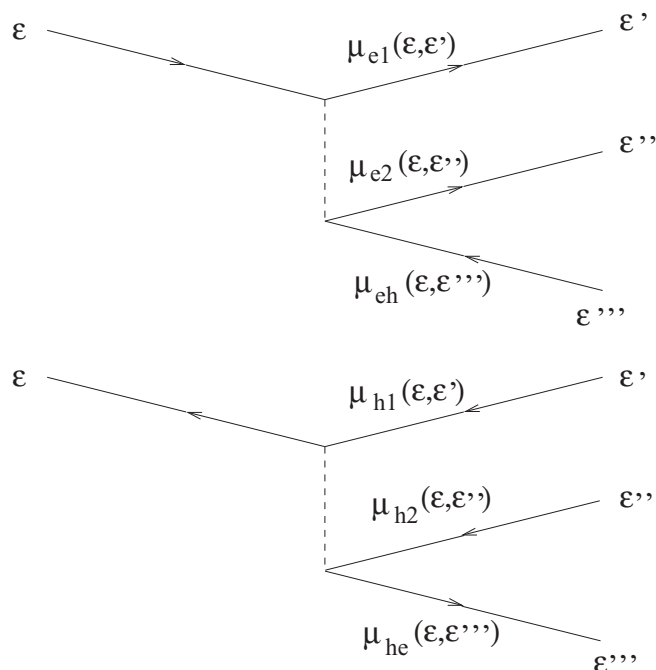


FIG. 5. Schematic illustration of cascade processes in which either a hot electron or a hot hole loses energy by creating a new electron-hole pair.

$S_h$  account for charge carriers injected as a result of tunneling processes. Of the remaining terms in the right hand side, the first one describes scattering out from a particular electron or hole energy and the rest of the terms describe scattering in of electrons and holes to the same energy. Figure 5 illustrates the different scattering processes in which either an electron (top) or hole (bottom) loses energy and creates an additional electron-hole pair, with the associated rates  $\mu$ , introduced in the work of Ritchie, used as labels:  $\mu_{e1}(\varepsilon, \varepsilon')$  is the rate (per unit time and energy) at which the primary electron is transferred to a state with energy  $\varepsilon'$ ;  $\mu_{e2}(\varepsilon, \varepsilon')$  is the rate at which a secondary hot electron with energy  $\varepsilon'$  is created in this process;  $\mu_{eh}(\varepsilon, \varepsilon')$ , finally, is the rate at which a secondary hot hole with energy  $\varepsilon'$  is created in the process. In the same way  $\mu_{h1}$ ,  $\mu_{h2}$ , and  $\mu_{he}$  refer to a process where a hot hole loses energy. Based on phase-space arguments one arrives at the following approximate expressions for the rates in a free-electron model,

$$\mu_{e1}(\varepsilon, \varepsilon') = \mu_{e2}(\varepsilon, \varepsilon') = \frac{\mu_0}{E_F} \frac{\varepsilon - \varepsilon'}{\varepsilon} \theta(\varepsilon - \varepsilon') \quad (\text{A3})$$

$$\mu_{eh}(\varepsilon, \varepsilon') = \frac{\mu_0}{E_F} \frac{\varepsilon + \varepsilon' - 2E_F}{\varepsilon} \theta(\varepsilon + \varepsilon' - 2E_F) \quad (\text{A4})$$

$$\mu_{h1}(\varepsilon, \varepsilon') = \mu_{h2}(\varepsilon, \varepsilon') = \frac{\mu_0}{E_F^2} (\varepsilon' - \varepsilon) \theta(\varepsilon' - \varepsilon) \quad (\text{A5})$$

$$\mu_{he}(\varepsilon, \varepsilon') = \frac{\mu_0}{E_F^2} (2E_F - \varepsilon - \varepsilon') \theta(2E_F - \varepsilon - \varepsilon'). \quad (\text{A6})$$

In these expressions  $\theta$  denotes a step function and all energies are measured relative to the band bottom.

We should stress here that we deal with transition rates per unit time, whereas in Ref. 31 transition rates per unit

length were used. We make the connection with that work by determining the overall rate  $\mu_0$  as

$$\mu_0 = v_F/\ell_0, \quad (\text{A7})$$

where the length  $\ell_0$  in a free-electron model is given by<sup>31</sup>

$$\frac{1}{\ell_0} = \frac{1}{18\gamma a_0} \left[ \arctan(1/\gamma) + \frac{\gamma}{1+\gamma^2} \right] \quad (\text{A8})$$

with

$$\gamma = \left( \frac{4}{9\pi^4} \right)^{1/6} r_s^{1/2}, \quad (\text{A9})$$

and  $a_0$  is the Bohr radius. For gold, with  $r_s \approx 3$ , we arrive at  $\ell_0 \approx 0.5$  nm, and we have used this value for all the calculations even though the STM tip below the Au coating consists of tungsten. The total scattering out rates for electrons and holes of energy  $\varepsilon$ ,  $1/\tau_e$  and  $1/\tau_h$ , are given by

$$\frac{1}{\tau_{e(h)}(\varepsilon)} = \int_{E_F}^{\varepsilon} \mu_{e(h)1}(\varepsilon, \varepsilon') d\varepsilon', \quad (\text{A10})$$

which yields inelastic mean free paths of the order 10 nm (and inelastic life times of the order of 10 fs) for excitations with an energy of 2 eV away from the Fermi level.

The left hand side of the diffusion equations contain a phenomenological damping term,  $u/\tau_0$  or  $v/\tau_0$ , which we will discuss further below, and a diffusion term with a diffusion constant

$$D = \frac{v_F \ell}{3}, \quad (\text{A11})$$

where  $\ell$  is a diffusion mean free path and  $v_F$  is the Fermi velocity. Here we have used  $\ell = 2$  nm, assuming a high degree of disorder in the tip.

To summarize, the model combines a local-in-space electron-hole cascade and diffusion without energy changes for the electrons and holes. The model is of course an approximation in that the typical energy transfer involves particles separated by a distance of the order of a screening length in the electron gas.

We treat electron and hole diffusion in a generic model geometry of a hemisphere of radius  $R_h = 50$  nm in each electrode. The primary hot electrons and holes are injected at the hemisphere centers, at  $r = 0$ , corresponding to the position of the tunneling contact. Given this model geometry, the electron and hole concentrations  $u$  and  $v$ , as well as the source distributions  $S_e$  and  $S_h$ , can be expanded in terms of spherical Bessel functions. For example, for electrons

$$u(\vec{r}, \varepsilon) = \sum_{n=1}^{\infty} u_n(\varepsilon) j_0(k_n r), \quad (\text{A12})$$

with corresponding expansion coefficients  $v_n(\varepsilon)$ ,  $S_{n,e}(\varepsilon)$ , and  $S_{n,h}(\varepsilon)$  for the other functions. We assume that electrons and holes are specularly reflected off the hemisphere's boundaries meaning that the wave numbers  $k_n$  must satisfy  $\partial j_0(k_n r)/\partial r = 0$  at  $r = R_h$  which yields the first few values for  $k_n$ ,

$$k_1 = 0, \quad k_2 \approx 4.49/R_h, \quad \text{etc.} \quad (\text{A13})$$

Inserting the expansions into Eqs. (A1) and (A2), we can solve for  $u_n$  and  $v_n$ , and get

$$u_n(\varepsilon) = \frac{S_{ne}^{(\text{tot})}(\varepsilon)}{1/\tau_0 + 1/\tau_e(\varepsilon) + Dk_n^2}, \quad (\text{A14})$$

and

$$v_n(\varepsilon) = \frac{S_{nh}^{(\text{tot})}(\varepsilon)}{1/\tau_0 + 1/\tau_h(\varepsilon) + Dk_n^2}. \quad (\text{A15})$$

The total source strengths  $S_{ne}^{(\text{tot})}$  and  $S_{nh}^{(\text{tot})}$  for hot electrons and holes of energy  $\varepsilon$  including both primary and secondary excitations are

$$S_{ne}^{(\text{tot})}(\varepsilon) = S_{ne}(\varepsilon) + \int_{E_{\min}}^{2E_F - \varepsilon} \mu_{he}(\varepsilon', \varepsilon) v_n(\varepsilon') d\varepsilon' + \int_{\varepsilon}^{E_{\max}} [\mu_{e1}(\varepsilon', \varepsilon) + \mu_{e2}(\varepsilon', \varepsilon)] u_n(\varepsilon') d\varepsilon' \quad (\text{A16})$$

and

$$S_{nh}^{(\text{tot})}(\varepsilon) = S_{nh}(\varepsilon) + \int_{2E_F - \varepsilon}^{E_{\max}} \mu_{eh}(\varepsilon', \varepsilon) u_n(\varepsilon') d\varepsilon' + \int_{E_{\min}}^{\varepsilon} [\mu_{h1}(\varepsilon', \varepsilon) + \mu_{h2}(\varepsilon', \varepsilon)] v_n(\varepsilon') d\varepsilon'. \quad (\text{A17})$$

The fact that the source strengths only involve the concentrations of ‘‘hotter’’ electrons and holes than  $\varepsilon$ , simplifies the solution. We can start by solving for electron and hole concentrations at the energies furthest from the Fermi energy and work inwards towards  $E_F$ .

The phenomenological damping time  $\tau_0$  is introduced to avoid the buildup of high artificial concentrations of hot holes and electrons for energies near the Fermi level because the model assumes specular reflection at  $r = R_h$ , meaning that the diffusion rate  $Dk_n^2$  in the denominators of Eqs. (A14) and (A15) is zero. To take into account that in reality holes and electrons can diffuse to regions far from the tunneling contact, the rate  $1/\tau_0$  is set to a nonzero value. We used  $1/\tau_0 \approx 2.7 \times 10^{12} \text{ s}^{-1}$  in the calculations here. The exact value of this rate has little influence on the hot hole and hot electron concentrations except in the immediate vicinity of the Fermi level.

The discussion has so far been general enough not to distinguish between the tip and sample, however, to calculate the occupation numbers  $f_e$  and  $f_c$  appearing in the main text we have to make the connection to the hot electron concentration  $u_e$  in the emitter electrode and the hot hole concentration  $v_c$  in the collector electrode, respectively. We do this by using the concentrations of hot electrons and holes right at the tunneling contact, i.e., at  $r = 0$  in the diffusion model, and relating them to the density of electron states  $D(E)$  at the energy in question. Provided energies are measured relative to the bottom of the free electron band we have

$$D(E) = \frac{m^{3/2} \sqrt{E}}{\pi^2 \hbar^3}. \quad (\text{A18})$$

We then get

$$f_e(E) = u_e(E, r = 0)/D(E), \quad (\text{A19})$$

and

$$1 - f_c(E) = v_c(E, r = 0)/D(E). \quad (\text{A20})$$

- 
- <sup>1</sup>T. N. Todorov, *Philos. Mag. B* **77**, 965 (1998).
- <sup>2</sup>N. Agraït, C. Untiedt, G. Rubio-Bollinger, and S. Vieira, *Phys. Rev. Lett.* **88**, 216803 (2002).
- <sup>3</sup>Y. C. Chen, M. Zwolak, and M. Di Ventra, *Nano Lett.* **3**, 1691 (2003).
- <sup>4</sup>G. Schulze, K. J. Franke, A. Gagliardi, G. Romano, C. S. Lin, A. L. Rosa, T. A. Niehaus, T. Frauenheim, A. Di Carlo, A. Pecchia, and J. I. Pascual, *Phys. Rev. Lett.* **100**, 136801 (2008).
- <sup>5</sup>N. Néel, J. Kröger, and R. Berndt, *Appl. Phys. Lett.* **95**, 203103 (2009).
- <sup>6</sup>A. M. Kadigrobov, R. I. Shekhter, S. I. Kulinich, M. Jonson, O. P. Balkashin, V. V. Fisun, Y. G. Naidyuk, I. K. Yanson, S. Andersson, and V. Korenivski, *New J. Phys.* **13**, 023007 (2011).
- <sup>7</sup>H. Pothier, S. Guéron, N. O. Birge, D. Esteve, and M. H. Devoret, *Phys. Rev. Lett.* **79**, 3490 (1997).
- <sup>8</sup>Y.-F. Chen, T. Dirks, G. Al-Zoubi, N. O. Birge, and N. Mason, *Phys. Rev. Lett.* **102**, 036804 (2009).
- <sup>9</sup>D. Hoffmann, J. Seifritz, B. Weyers, and R. Möller, *J. Electron Spectrosc. Relat. Phenom.* **109**, 117 (2000).
- <sup>10</sup>G. Hoffmann, R. Berndt, and P. Johansson, *Phys. Rev. Lett.* **90**, 046803 (2003).
- <sup>11</sup>G. Schull, N. Néel, P. Johansson, and R. Berndt, *Phys. Rev. Lett.* **102**, 057401 (2009).
- <sup>12</sup>The tunneling of electrons with elevated energies has been invoked in an early report on the emission of light from metal particle films (Ref. 13).
- <sup>13</sup>P. G. Borzjak, O. G. Sarbej, and R. D. Fedorowitsch, *Phys. Status Solidi* **8**, 55 (1965).
- <sup>14</sup>P. Johansson, R. Monreal, and P. Apell, *Phys. Rev. B* **42**, 9210 (1990).
- <sup>15</sup>N. L. Schneider, G. Schull, and R. Berndt, *Phys. Rev. Lett.* **105**, 026601 (2010).
- <sup>16</sup>R. Pechou, R. Coratger, F. Ajustron, and J. Beauvillain, *Appl. Phys. Lett.* **72**, 671 (1998).
- <sup>17</sup>A. Downes, P. Dumas, and M. E. Welland, *Appl. Phys. Lett.* **81**, 1252 (2002).
- <sup>18</sup>P. Johansson, *Phys. Rev. B* **58**, 10823 (1998).
- <sup>19</sup>T. Uemura, M. Furumoto, T. Nakano, M. Akai-Kasaya, A. Salto, M. Aono, and Y. Kuwahara, *Chem. Phys. Lett.* **448**, 232 (2007).
- <sup>20</sup>Z.-C. Dong, X. L. Zhang, H. Y. Gao, Y. Luo, C. Zhang, L. G. Chen, R. Zhang, X. Tao, Y. Zhang, J. L. Yang, and J. G. Hou, *Nat. Photonics* **4**, 50 (2010).
- <sup>21</sup>A. Fujiki, Y. Miyake, Y. Oshikane, M. Akai-Kasaya, A. Saito, and Y. Kuwahara, *Nanoscale Res. Lett.* **6**, 347 (2011).
- <sup>22</sup>W. S. Fann, R. Storz, H. W. K. Tom, and J. Bokor, *Phys. Rev. B* **46**, 13592 (1992).
- <sup>23</sup>W. S. Fann, R. Storz, H. W. K. Tom, and J. Bokor, *Phys. Rev. Lett.* **68**, 2834 (1992).
- <sup>24</sup>E. Knoesel, A. Hotzel, and M. Wolf, *Phys. Rev. B* **57**, 12812 (1998).
- <sup>25</sup>G. Hoffmann, L. Libioulle, and R. Berndt, *Phys. Rev. B* **65**, 212107 (2002).
- <sup>26</sup>J. Aizpurua, S. P. Apell, and R. Berndt, *Phys. Rev. B* **62**, 2065 (2000).
- <sup>27</sup>G. Hoffmann, J. Aizpurua, P. Apell, and R. Berndt, *Surf. Sci.* **482**, 1159 (2001).
- <sup>28</sup>K. Meguro, K. Sakamoto, R. Arafune, M. Satoh, and S. Ushioda, *Phys. Rev. B* **65**, 165405 (2002).
- <sup>29</sup>N. Agraït, A. L. Yeyati, and J. M. van Ruitenbeek, *Phys. Rep.* **377**, 81 (2003).
- <sup>30</sup>N. L. Schneider, F. Matino, G. Schull, S. Gabutti, M. Mayor, and R. Berndt, *Phys. Rev. B* **84**, 153403 (2011).
- <sup>31</sup>R. H. Ritchie, *J. Appl. Phys.* **37**, 2276 (1966).