Research article

Jacob B. Khurgin*

Fundamental limits of hot carrier injection from metal in nanoplasmonics

https://doi.org/10.1515/nanoph-2019-0396
Received September 30, 2019; revised October 27, 2019; accepted October 29, 2019

Abstract: The evolution of non-equilibrium carriers excited in the process of decay of surface plasmon polaritons (SPPs) in metal is described for each step – from the generation of carriers to their extraction from the metal. The relative importance of various carrier-generating mechanisms is discussed. It is shown that both the generation of carriers and their decay are inherently quantum processes as, for realistic illumination conditions, no more than a single SPP per nanoparticle exists at a given time. As a result, the distribution of non-equilibrium carriers cannot be described by a single temperature. It is also shown that the originally excited carriers that have not undergone a single electron-electron scattering event are practically the only ones that contribute to the injection. The role of momentum conservation in carrier extraction is discussed, and it is shown that, if all the momentum conservation rules are relaxed, it is the density of states in the semiconductor/dielectric that determines the ultimate injection efficiency. A set of recommendations aimed at improving the efficiency of plasmonic-assisted photodetection and (to a lesser degree) photocatalysis is made in the end.

Keywords: plasmons; hot carriers; photodetectors.

1 Introduction

The last two decades have seen a vigorous growth of exploration in plasmonics [1–3] and, in a broader sense, in the interaction between light and free carriers in metal or other media (such as doped semiconductors). The salient feature of metallic structures is their ability to concentrate optical fields into small volumes that are not limited by diffraction. A large number of plasmonic devices with enhanced performance in various regions of electromagnetic spectra have been conceived [4] and, to a certain degree, demonstrated since the turn of the millennium, including sources [5], detectors [6, 7] and modulators [8, 9] of radiation, as well as a wide range of sensors. Yet, with the exception of sensors [10, 11], plasmonic devices have failed to enter the mainstream for one reason – a very high loss inherent in all metals. This loss is innate to metals because a large fraction of energy is contained in the kinetic motion of carriers, which scatter energy at the rate of up to 10^{14} s^{-1} or even more [12]. Alternatively, if one wants to put it in a "quantum mindset," the aforementioned large loss can be explained by the large density of both occupied and empty states below and above Fermi energy, respectively, and, hence, the large rate of transitions between those states as postulated by the Fermi golden rule. The loss is particularly high at shorter wavelengths, while at longer wavelengths the damage from loss is less extensive [13–15] as the electromagnetic field does not really penetrate the metal. Thus, the functional metal-based devices in the mid-infrared (IR) and terahertz ranges are not really plasmonic in the correct sense of that word.

It is the realization that loss in plasmonics is inevitable that has prompted a significant part of the plasmonic community to re-examine the issue and shift the focus of their efforts from the futile battle to reduce the absorption in metals to the quest for creative use of that absorption [16, 17], which, in fact, should not be thought of as an irretrievable loss but rather as the transfer of energy from plasmons first to the single particle excitations in metal (electron-hole pairs) and then to the lattice vibrations. If the absorbed energy can be recaptured on one of the stages before equilibrium with the surroundings is achieved, it can be put to productive use, as has been indeed demonstrated by a number of groups [6, 18–20]. The stage at which absorbed energy can be captured with the least effort is obviously just after energy has been transferred to the lattice (which, in general, is far less than a picosecond). Depending on how well the plasmonic entities are isolated from

***Corresponding author: Jacob B. Khurgin,** Johns Hopkins University, Baltimore MD 21218, USA, e-mail: jakek@jhu.edu. https://orcid.org/0000-0003-0725-8736

the surroundings, they can be heated by hundreds of degrees, and that rise in temperature can be used for diverse applications ranging from cancer therapy [21] to thermophotovoltaics [22].

Capturing energy at the first stage, when it has just been transferred from the plasmon polaritons to the electron-hole pairs and before it has moved down the line to the lattice phonons, is obviously significantly more difficult because it has to be done on the sub-picosecond scale, but is also potentially far more rewarding as the kinetic energies (relative to the Fermi energy) of these so-called "hot carriers" are commensurate with the photon energy, i.e. they correspond to tens of thousands of degrees of Kelvin, rather beyond the melting point of the metals. For this reason, the hot carriers have sufficient energy to do what for the carriers in equilibrium with the lattice is impossible, even if the lattice is heated almost to the point of melting. In particular, hot carriers may have energy sufficient to overcome the binding force that keeps them inside the metal and carry charge into the adjacent material, which can be semiconductor, dielectric, or a solution surrounding the metal nanoparticle. Once the barrier is surmounted, the charges are separated, and, by collecting the charges at contacts, one can operate the device as a photodetector or a photovoltaic cell, as has indeed been done in many works. Alternatively, the charges can serve as catalysts for various chemical reactions that take place either directly on the metal surface [23–25], or, more often, after first being transferred to a dielectric or semiconductor material, such as TiO_2 [26, 27].

At this time, a large body of work has been accumulated on the process of charge transfer of hot carriers from plasmonic metals into semiconductors or dielectrics [27–39]. The results (in terms of the injection efficiency) have varied among the different groups, sometimes by orders of magnitude and more. The confusion has not been greatly helped by numerous theories that have been developed to explain the observed phenomena. Trying not to sound too critical of all the worthwhile efforts expended in the attempt to make sense of hot carrier injection, we would nevertheless mention the common trend of overreliance on numerical methods in many of the prior works, to the detriment of physical understanding. Moreover, recently, the whole concept of carrier injection has been put into question, and an attempt has been made to explain many experimental data simply via the heating of nanoparticles (i.e. what we referred to before as a second stage of energy transfer) [40, 41]. As shown later, this valiant endeavor, while not being entirely correct, is definitely not without a merit.

Faced with this reign of confusion in the field, I undertake this modest effort to shed some light onto hot carrier genesis, exodus, and decay in plasmonic metals. As I already mentioned, this is not an homage to all the prior works, nor is this their explicit critique, so readers glancing through these pages with the sole goal of establishing that a particular prior work has been mentioned (no matter in what context) will be disappointed and should seek and find solace in multiple articles proliferating elsewhere. In my opinion, referring the reader to a few solid review articles [42, 43] in addition to all that have been mentioned above is sufficient to establish the framework for the present work. I reiterate that the goals of this work are very modest – to present what I believe is a simple physical picture of hot carrier generation, decay, and emission and to establish what can be a maximum injection efficiency for a given metal/semiconductor (dielectric) combination.

The work is structured in the following way. Section 2, understandably, if ambitiously, entitled "Genesis," is essentially a very short review of my prior work [44] that has established how hot carriers are generated by different mechanisms and what their distribution is in energy, space, and angular coordinates. However, also in this section, an often-overlooked feature in plasmonics is uncovered – the fact that under most practical conditions, only a single surface plasmon polariton (SPP) is excited on a given nanoparticle. In Section 3, which, in line with the Pentateuch is called "Numbers," I make a key distinction between the "ballistic" or "first-generation" carriers produced by light and all the subsequent "generations" spawned by fast electron-electron (EE) scattering, with progressively lower energies, all the way to the thermalized hot carriers that can be characterized by the electron temperature T_e. I show that for the most practical situations, it is only that first "generation" of ballistic carriers that stands a decent chance of surmounting the barrier keeping them inside the metal. In Section 4, logically called "Exodus," I consider the role of lateral momentum conservation in the electron transfer across the barrier and show that, if momentum conservation is fully relaxed by interface roughness, the injection efficiency can be greatly increased and depends only on the density of states (DOS) in two adjacent media. It is this DOS ratio that determines the ultimate limit for electron injection from the metal, in total analogy to the "$4n^2$" limit for the light capture in dielectric determined by Yablonovitch almost four decades ago [45, 46]. Section 5 is unsurprisingly devoted to the conclusions that hopefully will be of some use to the community.

2 Genesis: hot carriers are generated. How and where?

2.1 How many SPPs are excited simultaneously?

Before reviewing the mechanisms leading to SPP decay into electron-hole pairs, it is instructive to roughly ascertain the numbers (for localized) and densities (for propagating) of SPPs involved in this process. One can start with the localized SPPs as shown in Figure 1A, in which the electric field E_{SPP} is enhanced relative to the incident field E_{IN} by roughly a factor of $F \approx Q = \omega/\gamma$ [1, 47], where γ is the total SPP decay rate (radiative and non-radiative) to be evaluated later. This ratio can be higher or lower depending on the geometry and, if resonant nano-antennas are involved, the ratio can be as high as $F \sim Q^2$ [48, 49]. Therefore, the energy density in the SPP mode can be found as $u_{SPP} = F^2 I_{IN} n/c$, where I_{IN} is the power density of the incoming radiation and n is the index of refraction surrounding the nanoparticle. The total energy of the SPP mode is then $U_{SPP} = F^2 I_{IN} n/c \times V_{SPP}$, where V_{SPP} is a volume of the SPP mode, commensurate with the size of nanoparticle. Consider then a spherical nanoparticle of 50 nm diameter and the light with average wavelength of 500 nm and refractive index $n = 1$. Then, the average number of SPPs in the mode is $N_{SPP} = U_{SPP}/\hbar\omega \approx 2 \times 10^{-9}\ F^2\ I_{IN}$, where the incident power density is in W/cm^2. Now, even if we assume a really high field enhancement factor $F^2 = 10^3$, one can see that in order to have on average one SPP at a given time, the incoming radiation should have power density of 1 MW/cm^2 in a spectral region around SPP resonance. For nanoparticles that are <50 nm in diameter, the required power density is even higher than 1 MW/cm^2 and reaches an extremely high value (100 MW/cm^2) for a 10-nm nanoparticle. For comparison, solar irradiance near the equator (1 sun) is only $I_{IN} \sim 0.1$ W/cm^2, so even concentrating this power by a factor 10^3 would still leave us with far less than 1 SPP at any given time. The situation is shown schematically in Figure 2A, where among many nanoparticles only a single one marked as "A" supports an SPP. After a plasmon lifetime $\tau_{SPP} = \gamma^{-1}$ elapses, there are no SPPs left within the volume as shown in Figure 2B, while the SPP energy is transferred to hot carriers in the marked nanoparticle "A." Then, as shown in Figure 2C, after another τ_{SPP} (although it can take significantly longer), an SPP is excited on the nanoparticle "B," and so on. Obviously, in photodetectors, the irradiance is typically far less than 1 MW/cm^2 as well. In other words, in practical situations, a given nanoparticle has no SPP in it most of the time – a single SPP is excited on rare occasions and then very quickly decays. This indicates that SPP decay is, by definition, a quantum process that cannot be described classically, as only one SPP is present at any given time. Another way to look at it is to establish the average interval between SPP excitations of a given nanoparticle as $\Delta\tau = \tau_{SPP}/N_{SPP}$, which, for the 50-nm nanoparticle with $\tau_{SPP} = 10$ fs considered above, $\Delta\tau = 0.5\ \mu s/I_{IN}$. Thus, it takes microseconds for the same nanoparticle to get excited again. Hence, the situation depicted in Figure 1A, in which the SPP mode is excited while some electron-hole pairs are present in the nanoparticle, is not realistic – at a given time, there may be an SPP excited on a nanoparticle or a hot electron-hole pair in it, or, most probably no excitation at all.

In the other relevant geometry (Figure 1B), if we consider a plasmonic or hybrid waveguide in which $P_{IN} = 1$ μW of power (which is a lot for the detector) propagates

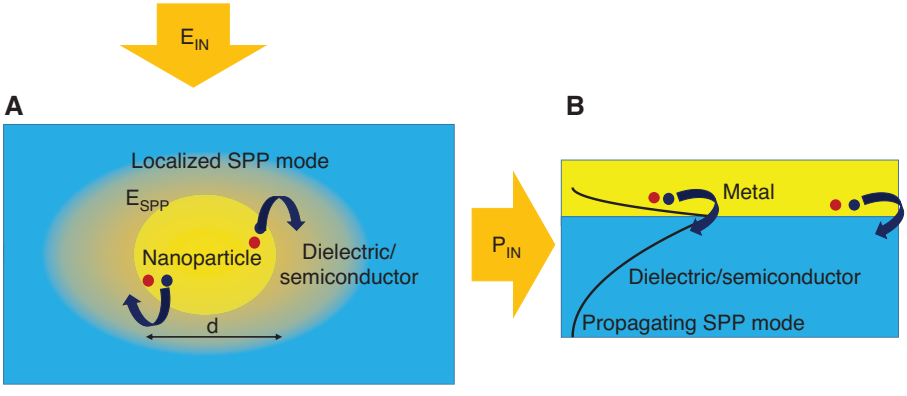

Figure 1: Plasmon-assited hot carrier generation and injection.
Geometry of hot electron-hole pair generation in the metal with subsequent injection into the semiconductor/dielectric using (A) localized SPPs (mostly used in photocatalysis) and (B) propagating SPPs (mostly used in photodetection). As explained in the text, at realistic optical powers, SPPs never coexist with hot carriers at the same time.

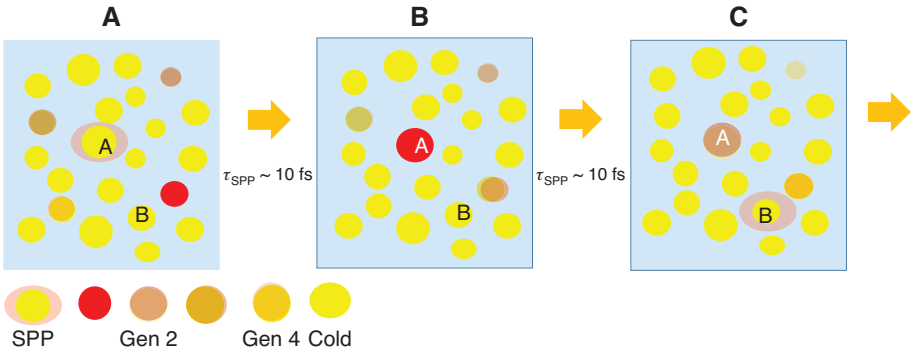

Figure 2: Dynamics of hot carrier generation and decay in the array of metal nanoparticles under typical photoexcitation conditions. At each point in time, only very few SPPs are excited and also very few nanoparticles contain non-equilibrium carriers. For example, in "frame" (A), only one nanoparticle A is excited with SPP and few other nanoparticles tinted red contain non-equilibrium carriers generated by previously excited and decayed SPPs, while a majority of nanoparticles tinted yellow remain "cold" or unexcited. In frame (B), the SPP on nanoparticle A has decayed into a hot electron-hole pair and no SPPs are present. In frame (C), a new SPP is excited on nanoparticle B while the existing non-equilibrium carriers continue to further cool down.

while being absorbed in metal, the linear density of the propagating SPPs is $dN_{SPP}/dl = P_{IN}n/c\hbar\omega \approx 0.01/\mu m$. As, typically, absorption takes place over a distance of just a few micrometers, once again the average number of SPPs inside the plasmonic waveguide at a given time is less than one, and, once again the process of hot carrier excitation should be treated as a quantum process.

It should be noted that in many experimental works, SPP excitation and hot carrier generation and decay have been studied using femtosecond lasers [29, 50–52], in which power densities can be much higher than 1 MW/cm^2. These time-resolved measurements do provide information about the evolution of hot carriers. However, they do it for the situation of multiple hot carrier excitations, which actually almost never occur in practical applications. For instance, when multiple SPPs are generated on the same nanoparticle, there is always a possibility of exciting the carriers with energy exceeding $\hbar\omega$; however, when the interval between SPP generations is long, such an event cannot realistically take place.

2.2 Four excitation mechanisms

We now review the four mechanisms that lead to the decay of SPP [44]. The first to be mentioned is the direct (i.e. no phonons or impurities are involved) interband absorption between the inner (4d or 5d) and outer (5s or 6s) shells of noble metals, as shown in Figure 3A. The energy gap separating the d shell and Fermi level residing in s-shell E_{ds} is equal to 2 eV for Au and 3 eV for Ag; therefore, the kinetic energy of the electron generated in the s-band (relative to the Fermi level) is only $E_{ib} < \hbar\omega - (E_F - E_d)$. Hence, only ultraviolet excitation can create carriers energetic enough to surpass the barrier Φ that is on the order of 0.5–1 eV. This is shown in Figure 4A, where the probability of energy distribution is $F_{hot,ib}(E) = 1/(\hbar\omega - E_{ds})$. As far as the angular distribution of the non-equilibrium carriers generated via interband absorption is concerned, the interband matrix element of momentum does not include wavevector k; hence, the angular distribution is uniform, $R_{ib}(\theta) = 1/2$, as shown in Figure 5A.

As for the holes generated in the d-shell, they have large effective mass. Therefore, they may have large potential energy relative to the Fermi level, but their kinetic energy is small, and, more important, their ballistic velocity is very low and the mean free path is short. The holes generated inside the metal have a very slim chance to reach the surface. Only when the photoexcited (first-generation) hole decays into three new second-generation particles (two holes and one electron) in s-shell does injection become possible; however, due to their reduced energies, for second-generation carriers (see below), the probability of such emission is rather low. For this reason, interband absorption only reduces the efficiency of hot carriers.

All other mechanisms are intraband, i.e. they involve absorption between two states with different wavevectors within the same s-band. This wavevector (momentum) mismatch needs to be somehow compensated. In the second mechanism, the compensation is provided by either a phonon or an impurity (defect) with wavevector $\boldsymbol{q}$, as shown in Figure 3B. As a result, SPP is absorbed and a hot electron and hot hole, each with an average energy of $\hbar\omega/2$, are generated. The energy distribution of the "first generation" of hot carriers is $F_{hot,ph}(E) = 1/\hbar\omega$,

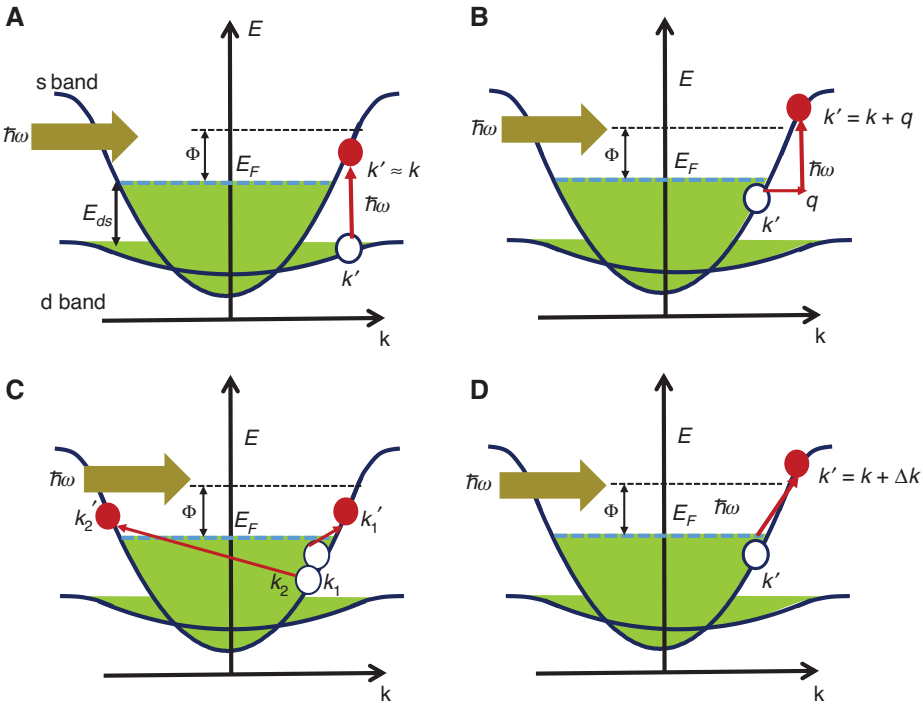

Figure 3: Four mechanisms of electron-hole pair generation in metals where the average kinetic energy of electrons and holes is low. (A) direct "vertical" interband transition; (B) phonon (or impurity)-assisted transition where the average energy of electrons or holes is $\hbar\omega/2$; (C) EE Umklapp scattering-assisted transition with two electron-hole pairs generated for each SPP and their average kinetic energies being $\hbar\omega/4$. (D) Landau damping or surface collision-assisted "tilted" transition where the average energy of electrons or holes is $\hbar\omega/2$. As one can see, the carriers generated via processes (B) and (D) are far more likely to have energy sufficient to overcome the surface (Schottky) barrier Φ and be injected into the adjacent semiconductor or dielectric.

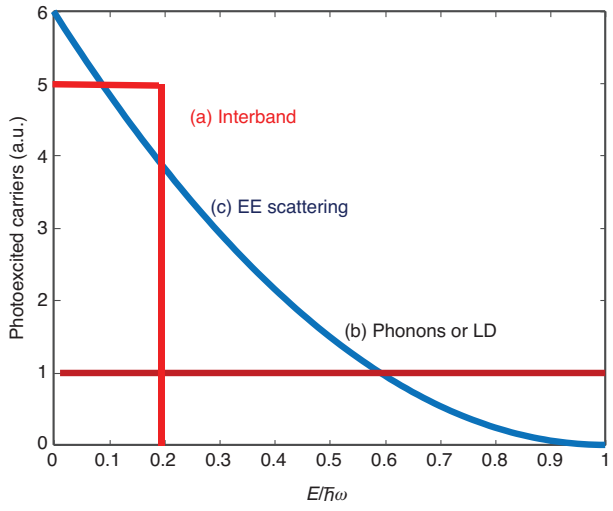

Figure 4: Energy distributions of carriers excited via different SPP decay mechanisms: (A) interband transitions; (B) phonon/defect-assisted transitions or Landau damping; (C) EE scattering-assisted transitions. Note that the area under curve (C) is twice as large as for the other two curves because two electron-hole pairs are generated for each decayed SPP.

where $E_F < E < E_F + \hbar\omega$ for electrons and $E_F > E > E_F - \hbar\omega$ for hot holes, as shown in Figure 4B. Conceptually, this process is not different from what is commonly referred to as "Drude" absorption, and the SPP damping rate due to this process can be found as $\gamma_{ph}(\omega) = \langle \tau_{ep}^{-1}(E) \rangle_E$, where the electron-phonon (or defect) scattering rate is averaged from $E_F - \hbar\omega$ to $E_F + \hbar\omega$ (while in the Drude formula, the scattering is taken on the Fermi level). The scattering rate for Ag is about $\gamma_{ph} \approx 3 \times 10^{13}\ s^{-1}$ and for Au is $\gamma_{ph} \approx 10^{14}\ s^{-1}$ [53]. However, it is important to stress that this is a quantum, not a classical, process. There is no so-called "classical" [32, 37] or "resistive" [31, 42] contribution to the absorption in which many carriers supposedly are instantly created. The energy of the absorbed SPP is almost entirely transferred to just two hot particles – an electron and a hole, and is not dissipated to a bath of multiple carriers near the Fermi level. That process takes much longer, as discussed in the next section. Still, some classical analogies remain true even in the quantum picture. As classically the carriers are accelerated along the direction of optical field, one would expect the photoexcited hot electrons and holes to preferentially travel in that direction. Indeed, detailed calculations show that photoexcited carriers have a normalized angular distribution:

$$R_{ph}(\theta) = \frac{3}{4}\cos^2\theta + \frac{1}{4}, \tag{1}$$

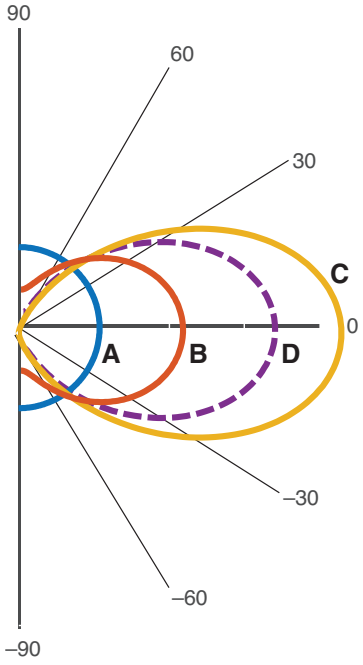

Figure 5: Angular distributions of carriers excited via different SPP decay mechanisms.
(A) interband transitions and EE scattering-assisted transitions; (B) phonon/defect-assisted transitions; (C) Landau damping; (D) effective distribution combining carriers generated by Landau damping on the surface and the carriers generated in the bulk by phonon/defect-assisted transitions that have reached the surface.

as shown in Figure 5, curve b. As the electric field in most plasmonic structures is close to being normal to the surface, the fraction of hot carriers going toward the surface is twice as large as for uniform distribution. One should also stress that while the phonon scattering rate at the Fermi level that determines Drude conductivity for low-frequency excitations goes down dramatically (proportionally to T^5 [54]) at low temperatures, for optical frequencies when photon energies significantly exceed the Debye energy, the phonon scattering rate is reduced by a relatively small amount and never goes to zero, as can be attested by experimental data in Refs. [55, 56].

The third mechanism by which SPP decays involves EE scattering [57, 58] (Figure 3C). In this process, two electrons and two holes share the energy of the decayed SPP, so on average the energy of each carrier is just $\hbar\omega/4$. Thus, calling them "hot" may not be correct and perhaps "warm" would be a better term. It is well known that, at low frequencies, the EE scattering contribution to the electrical resistance (and therefore ohmic loss) is negligibly small. The reason for this is the fact that the total momentum of carriers undergoing EE scattering is conserved, i.e. $\boldsymbol{k}_1' + \boldsymbol{k}_2' = \boldsymbol{k}_1 + \boldsymbol{k}_2$. As long as the band can be considered parabolic near the Fermi surface, the total current can be found as $\boldsymbol{I} = -e\boldsymbol{v}_1/l - e\boldsymbol{v}_2/l = -(e\hbar/lm)(\boldsymbol{k}_1 + \boldsymbol{k}_2)$, where l is the length, $\boldsymbol{v}$ is velocity, and m is the effective mass. Therefore, the total current is conserved and no energy is dissipated via EE scattering. However, for optical frequencies, the situation is dramatically different because the photon energy is sufficiently large to initiate the Umklapp processes [54, 59] in which one of the photoexcited electrons is promoted into the adjacent Brillouin zone, so that momentum conservation relation becomes $\boldsymbol{k}_1' + \boldsymbol{k}_2' = \boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{g}$, where $\boldsymbol{g}$ is the reciprocal lattice vector. Obviously, the velocity and current now change as a result of EE scattering and the SPP decays. The EE scattering-assisted SPP damping rate has been found as $\gamma_{ee}(\omega) = F_U(\omega)\tau_{ee}^{-1}(\omega)$, where the EE scattering rate is [60]

$$\tau_{ee}^{-1} \approx \frac{\pi}{24}\frac{E_F}{\hbar}\left(\frac{\hbar\omega}{E_F}\right)^2, \tag{2}$$

and $F_U(\omega)$ is the fraction of the total EE scattering events that are Umklapp processes. This fraction is typically on the order of 0.2–0.5. It follows that EE-assisted SPP decay becomes prominent at short wavelengths [61, 62] and for photon energies >2 eV $\gamma_{ee} \sim 10^{14}$ s^{-1}, i.e. at least as large as the phonon-assisted SPP damping rate. At the same time, for photon energies <1 eV, for example for the telecom range ($\hbar\omega = 0.8$ eV), EE-assisted damping is not important. The energy distribution of the first-generation carriers excited with the assistance of EE scattering is

$$F_{hot,ee} = 2 \times 3(\hbar\omega - E_F)^2/(\hbar\omega)^3, \tag{3}$$

where the factor of 2 indicates that two hot electrons are excited by each SPP decay event. This energy distribution is plotted in Figure 4C. As far as the angular distribution goes, due to the involvement of reciprocal lattice vectors, the distribution is roughly uniform, as shown in Figure 5A. For all three SPP decay mechanisms outlined thus far, the spatial distribution of non-equilibrium carrier generation simply follows the density of the SPP energy $\boldsymbol{\mathcal{E}}(\boldsymbol{r})^2$, as shown in Figure 6.

The fourth and last SPP decay channel (Figure 3D) is referred to either classically (or phenomenologically) as surface collision-assisted decay [63, 64] or in the quantum picture as Landau damping (LD) [65–68]. Classically, when an electron collides with the surface, momentum can be transferred between the electron and the entire metal lattice, in a way similar to what happens when an electron collides with a phonon or defect. This relaxes the momentum conservation rules and, as first done by Kreibig and Vollmer [69], one can simply introduce the

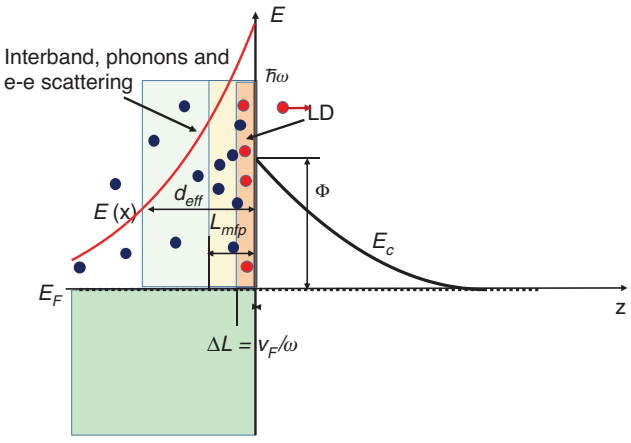

Figure 6: Energies and locations of non-equilibrium carriers generated via different SPP decay mechanisms.
The carriers generated via interband, phonon/defect-assisted, and EE scattering-assisted transitions within the skin depth (or for a small nanoparticle within the entire bulk), as shown by black circles. With Landau damping, the carriers (red circles) are generated with a layer near the surface $\Delta L = v_F/\omega$, which is much thinner than the skin depth.

surface collision rate $\gamma_{sc} \sim v_F/d$, where d is the size of a nanoparticle. Quantum mechanically, the absorption is the result of the spatial localization of optical field. As the field is localized, its Fourier transform contains all the spectral components, some of which are higher than $\Delta k = \omega/v_F$, where v_F is Fermi velocity, which, for Au and Ag, is about 1.4×10^8 cm/s. These spectral components provide necessary momentum matching, which allows absorption of SPP without assistance from the phonons or defects. This process is commonly referred to as LD [47, 70, 71] and is characterized by the existence of the imaginary part of the wavevector-dependent (non-local) dielectric permittivity of the metal described by Lindhard's formula [72]:

$$\varepsilon(\omega,k) = \varepsilon_b + \frac{3\omega_p^2}{k^2 v_F^2}\left[1 - \frac{\omega}{2kv_F}\ln\frac{\omega + kv_F}{\omega - kv_F}\right], \tag{4}$$

which obviously has an imaginary part for $|k| > \omega/v_F$. The rate of SPP decay due to LD is

$$\gamma_{\mathrm{LD}} = \frac{3}{8} v_F / d_{\mathrm{eff}}, \tag{5}$$

where

$$d_{eff} = \frac{\int_{\mathrm{metal}} \mathcal{E}(\mathrm{r})^2 \, \mathrm{d}V}{\int_{\substack{\mathrm{metal} \\ \mathrm{surface}}} \mathcal{E}_{\perp}^2(\mathrm{r}) \mathrm{d}S}, \tag{6}$$

is the volume-to-surface ratio of the mode in the metal and $\mathcal{E}_{\perp}(\boldsymbol{r})$ is the normal component of the electric field. Clearly both phenomenological and more exact full quantum treatments provide similar results – if one uses Eqs. (5) and (6) on a spherical nanoparticle, one obtains $\gamma_{LD} = 0.75 v_F/d$ instead of $\gamma_{sc} = v_F/d$.

It is important to note [44] that the hot carriers generated by surface collisions (LD) are all located within a thin layer of thickness $\Delta L = 2\pi/\Delta k = v_F/\nu$, where ν is an optical frequency shown in Figure 6. As one can see, ΔL is the distance covered by the electron over one optical period, e.g. for gold and 700 nm excitation, is only about 3 nm, which is obviously shorter than the mean free path of an electron between collisions (typically 10–20 nm). Therefore, one-half of the carriers excited via LD will always end up at the surface, which is one reason why LD is the most favorable mechanism of carrier generation for their ejection from the metal. The second reason is that the angular distribution of the carriers excited via LD is highly non-uniform:

$$R_{\mathrm{LD}}(\theta) \sim 2|\cos^3\theta|, \tag{7}$$

as shown in Figure 5C. As one can see, the fraction of carriers that impinge on the surface at normal incidence is increased by a factor of 4 compared to the uniform distribution and by a factor of 2 compared to the distribution of the carriers generated by phonon-assisted processes. It should be noted that, if the dimension of the nanoparticle becomes comparable to or less than ΔL, then the dielectric constant can no longer be defined and the "bulk" Lindhard's formula [Eq. (4)] is no longer valid. Then, one should revert to calculation of the discrete transitions between the confined levels in the metallic "quantum dot." However, once one takes into account the broad distribution of the sizes, the peaks of absorption are broadened and the results do not differ dramatically from the ones obtained in this work until one goes to clusters of <1 nm in size.

3 Numbers: hot carriers decay, but how fast?

3.1 Is there such thing as equilibrium electron temperature and how hot can it be?

We have established that SPP shown in Figure 7A decays within its lifetime $\tau_{\mathrm{SPP}} \sim \gamma^{-1}$, where the sum γ of all four SPP decay rates is outlined above as well as the radiative

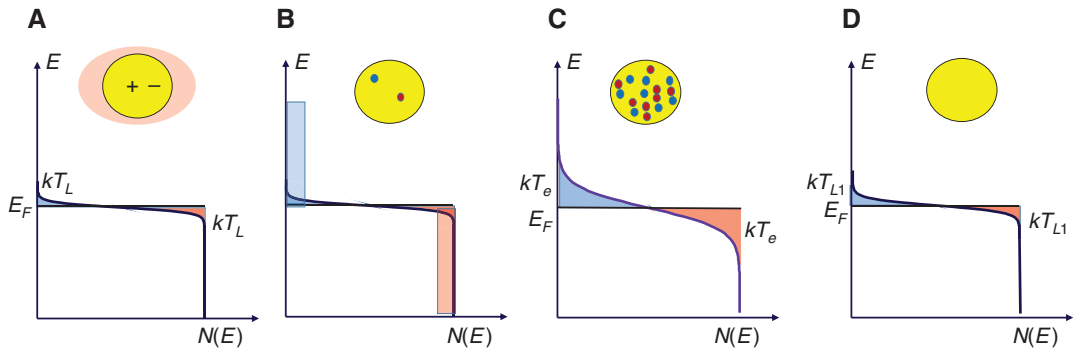

Figure 7: Conventionally assumed picture of hot carrier generation and relaxation in metal nanoparticles.
(A) An SPP is excited on a nanoparticle while the carriers inside are distributed according to Fermi-Dirac statistics with equilibrium lattice temperature T_L. (B) After the SPP lifetime τ_{SPP}, this SPP has decayed and, as a result, non-equilibrium electron-hole pairs with energies ranging from $E_F - \hbar\omega$ to $E_F + \hbar\omega$ have been generated. (C) After the electron thermalization time commensurate with τ_{ee}, all the electrons are distributed according to Fermi-Dirac statistics with electron temperature $T_e > T_L$. (D) After the electron-lattice thermalization time τ_{el}, the electron and lattice are at equilibrium with a new lattice temperature $T_{L1} > T_L$. This temperature will eventually decrease back to the equilibrium lattice temperature T_L after time τ_L determined by the nanoparticle environment.

decay rate γ_{rad} that is typically small in comparison. Once the first-generation non-equilibrium carriers have been excited as shown in Figure 7B, they decay via both EE- and phonon (or defect)-assisted processes. Conventionally, it is assumed that fast EE scattering $\tau_{ee} \sim 10$ fs [50, 51, 73–75], defined in Eq. (2), quickly establishes thermal equilibrium among the electrons with temperature T_e that is significantly higher than the equilibrium lattice temperature T_L (Figure 7C). Then, the electron temperature relaxes as the energy is transferred to the lattice with characteristic relaxation time τ_{el} that is a couple of orders of magnitude longer than τ_{ee}, as shown in Figure 7D. Often, this time is erroneously referred to as electron-phonon scattering time, but this is of course wrong – the electron phonon scattering time τ_{ep}, defined in the previous section, is roughly the same order of magnitude as τ_{ee} but it takes many scattering events to reduce the energy of hot electrons because, in each event, only a small amount of energy is being lost by an electron. It is for this reason that the first-generation carriers that underwent a phonon scattering event can be considered quasi-ballistic, as they largely keep their energy and their distribution in the momentum space does not change significantly.

The whole process, in general, cannot be characterized with two separate relaxation times, especially because τ_{ee} gradually increases as electrons lose energy, while τ_{el} gradually decreases as more secondary hot carriers are generated. We shall return to it later, but, at any rate, it is reasonable to assume that, for some time, all the energy absorbed by the metal as a result of SPP decay is stored in the electron gas, causing an effective rise of average temperature:

$$\overline{T}_e - T_L \sim \frac{\gamma_{nr}\tau_{el}}{C_{el}} u_{SPP} = \frac{\gamma_{nr}\tau_{el}}{C_{el}} F^2 I_{IN} n / c, \tag{8}$$

where the specific heat of electrons can roughly be estimated as

$$C_{el} = \frac{\pi^2}{2}\frac{k_B^2 T}{E_F} N_e \approx 0.025\, k_B N_e \approx 0.018\ \mathrm{J/K\,cm^3}, \tag{9}$$

where $E_F = 5.56$ eV (for Au) is Fermi energy and $N_e = 5.9 \times 10^{22}$ cm^{-3} (for Au) is the density of the free electrons. The specific heat is low because only about 1 out of 40 carriers residing within the $\pi^2 k_B T/4 \approx 64$ meV Fermi level participate in heat exchange. Substituting Eq. (9) into Eq. (8) and assuming a realistic value of $\gamma_{nr}\tau_{el} \sim 100$, we obtain

$$\overline{T}_e - T_L \sim 2 \times 10^{-7} F^2 I_{IN}\ \mathrm{K}, \tag{10}$$

where the input irradiance is once again in units of W/cm^2. With $F^2 = 10^3$ and input irradiance of 1000 suns (100 W/cm^2), the average electron temperature rise is miniscule at 0.02 K, in agreement with Ref. [76]. However, what about the instant temperature rise? According to our estimates, the average time interval between two SPP absorption events is typically longer than the "storage time" τ_{el}. The temperature rise following absorption of a single SPP is simply

$$\Delta T_{e,inst} = \hbar\omega / C_{el} V \approx 1.6 \times 10^4\ \mathrm{K\,nm^3/V}. \tag{11}$$

The instant increase is much higher than the average one and, for really small nanoparticles of 10 nm diameter,

it can be as high as 12 K, which is easy to understand as the energy of the first-generation electron-hole pair is eventually split between roughly 700 thermally active electrons residing near the Fermi level. However, even with this temperature increase (which in units of energy corresponds to about 1 meV), there would be no noticeable impact on carriers surmounting the energy barrier Φ that is typically hundreds of meV high. One should note that the result [Eq. (11)] is remarkable, as it points to the quantum nature of hot carrier generation with the equilibrium temperature of hot carriers depending only on the volume of nanoparticle and the photon energy, and totally independent on input power density. Furthermore, even if the excitation is continuous, the picture of all carriers settling at the same average electron temperature $\overline{T}_e$ is deeply flawed – the correct picture is that at any given time, a relatively small fraction of nanoparticles have electrons in them excited to some temperatures $\Delta T_{e,inst}$, which are different for different nanoparticles due to their different volumes, while the majority of nanoparticles experience no electron temperature increase at all, as shown Figure 1 where the vast majority of the nanoparticles remain "cold" at any given moment. Also, it is important to note that once the energy is transferred to the lattice, the temperature of the lattice T_{L1} is different from the equilibrium T_L and it takes some time τ_L for it to reach equilibrium. This time, τ_L is determined mostly by how efficient the energy transfer between the nanoparticle and the surrounding (or adjacent) medium is. If τ_L is longer than the mean interval between SPP excitations of the same nanoparticle $\Delta\tau$, the equilibrium lattice temperature T_L increases, resulting in thermionic emission over the barrier.

Overall, the intermediate conclusion here is that if one posits that the hot carriers have thermalized at some electron temperature T_e, these carriers definitely do not have sufficient energy to contribute to the injection into the semiconductor or the dielectric and, most probably, they cannot contribute to a chemical reaction on the metal surface either, as stated by Sivan et al. [40, 41, 76]. (The reason for the ambiguity here is that the nature of the energy barrier that the electron must surpass to cause a chemical reaction is not clearly defined.) Therefore, one must follow the decay of the excited hot carriers step by step in order to ascertain the probability of them being ejected out of the metal after each step, where the step is defined by a single collision of hot carriers with the thermal carriers (EE scattering) or with phonons and impurities.

3.2 Generations of non-equilibrium carriers

These relaxation steps are shown in Figure 8. On the left, the SPP decay causes the excitation of a single electron-hole pair – the first generation with energy $E_{1,n}$ and $\sum_{n=1}^{2} E_{1,n} = \hbar\omega$ (assuming that the hole energies are counted down from the Fermi level). Then, either an electron or a hole scatters off the electron residing below the Fermi level, thus creating another electron-hole pair. Once both an electron and a hole scatter once (which, on average, should take time τ_{ee}), there are three second-generation holes and electrons each with energies $E_{2,n}$ and $\sum_{n=1}^{6} E_{2,n} = \hbar\omega$. Thereafter, on average, another time interval

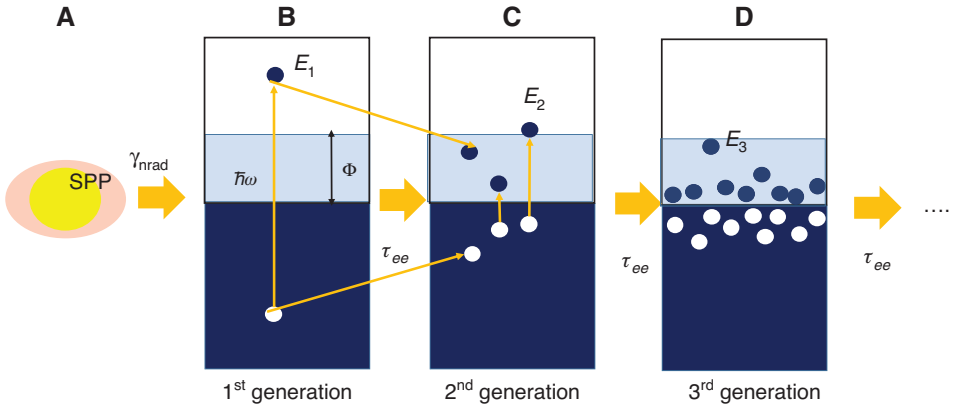

Figure 8: Quantum picture of carrier generation and relaxation in metal nanoparticles.
(A) An SPP is excited on a nanoparticle. (B) An SPP has decayed, engendering a primary (first-generation) electron-hole pair. (C) Each of the first-generation carriers decays into three second-generation carriers. (D) Each of the second-generation carriers decays into three third-generation carriers.

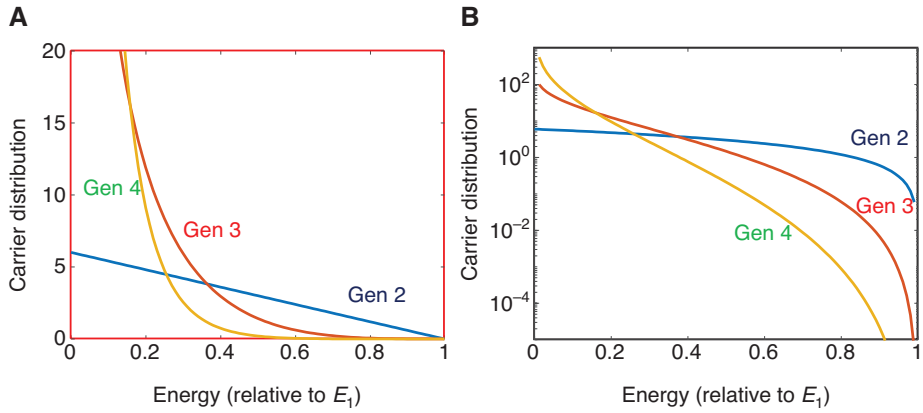

Figure 9: Energy distribution of the second through fourth generation of carriers generated as a result of decay of a single first-generation carrier with energy E_1.
(A) linear scale; (B) logarithmic scale.

τ_{ee} elapses (this interval may be longer than the original as the EE scattering is energy dependent), each of the second-generation carriers engenders three third-generation carriers with energies $E_{3,n}$ and $\sum_{n=1}^{18} E_{2,n} = \hbar\omega$. The process continues until the average energy of the M-th generation, $\langle E_{M,n}\rangle_n = \hbar\omega / 2\cdot 3^{M-1}$, becomes comparable to $k_B T$, so that there is no distinction between the "hot" and "cool" carriers; therefore, it takes roughly time

$$\tau_{e,cool} = (M-1)\tau_{ee} \approx \log_3(\hbar\omega / 2k_B T)\tau_{ee} \tag{12}$$

for the electrons to cool down to some kind of equilibrium between themselves. For SPP energies <1.5 eV, it takes no more than three scattering events to cool down the electrons; hence, $\tau_{e,\mathrm{cool}} < 4\tau_{ee}$.

Thus, while the cool-down time is of the same order of magnitude as the EE scattering time, it is definitely larger than it by a factor of a few. Therefore, the ubiquitous statement stubbornly permeating the literature that a single scattering event is sufficient to establish the equilibrium of the electrons [31, 32, 36, 37, 42] is incorrect. Obviously, during this time interval, there will be electron-phonon scattering events, because, remember, that τ_{ee} and τ_{ep} are roughly of the same order of magnitude. However, these events cause insignificant loss of energy for each hot carrier and thus can be safely disregarded.

Let us now consider the distribution of the second-through fourth-generation hot carriers in energy space. When the electron of the first-generation hot carriers with energy distribution $f_1(E) = \delta(E - E_1)$ decays into three new second-generation carriers, their distribution is

$$f_2(E, E_1) = 2(E_1 - E)/E_1^2. \tag{13}$$

Then, these carriers decay into nine third-generation carriers, whose distribution can be found as

$$f_3(E, E_1) = \int_E^{E_1} f_2(E_2, E_1) f_2(E, E_2)\,\mathrm{d}E_2 = \frac{4}{E_1^2}\left[(E_1 + E)\log\frac{E_1}{E_3} - 2(E_1 - E)\right], \tag{14}$$

and then into 27 fourth-generation carriers, with a distribution of

$$f_4(E, E_1) = \int_{E4}^{E_1} f_3(E_3, E_1) f_2(E, E_3)\,\mathrm{d}E_3 = \frac{4}{E_1^2}\left[(E_1 - E_4)\left(\log^2\frac{E}{E_4} + 12\right) - 6(E_1 + E_4)\log\frac{E}{E_4}\right]. \tag{15}$$

Note that even though the functions f_m for $m > 2$ diverge near zero energy, they are all perfectly integrable to $\int_0^{E_1} f_m(E, E_1)\,\mathrm{d}E = 1$. The distributions of the total number of carriers in each generation, $N_m(E, E_1) = 3^{m-1} f_m(E, E_1)$, are shown in Figure 9, with energies E_m scaled relative to energy E_1. As one can see, the distribution quickly shifts to lower energies; however, when plotted on log scale in Figure 9B, the curves are not linear and, therefore, one cannot ascribe a single electron temperature T_e to the carriers.

Next, we determine the distribution of all carriers generated by photons with energy $\hbar\omega$ as

$$f_m(E, \hbar\omega) = \int_{E_m}^{\hbar\omega} f_m(E, E_1)\,\mathrm{d}E_1 \tag{16}$$

and obtain (assuming the original distribution associated with phonon, defect-assisted, or LD process)

$$f_1(E,\hbar\omega)=F_{hot,ph}(E)=\frac{1}{\hbar\omega};\ E\le\hbar\omega$$
$$f_2(E,\hbar\omega)=\frac{2}{\hbar\omega}\left(\frac{E}{\hbar\omega}-1-\log\frac{E}{\hbar\omega}\right)$$
$$f_3(E,\hbar\omega)=\frac{2}{\hbar\omega}\left(\log^2\left(\frac{E}{\hbar\omega}\right)+2\log\frac{E}{\hbar\omega}\left(\frac{E}{\hbar\omega}+2\right)+6\left(1-\frac{E}{\hbar\omega}\right)\right)$$
$$f_4(E,\hbar\omega)=\frac{4}{\hbar\omega}\left(-\frac{1}{3}\log^3\left(\frac{E}{\hbar\omega}\right)+\log^2\left(\frac{E}{\hbar\omega}\right)\left(\frac{E}{\hbar\omega}-3\right)-4\log\frac{E}{\hbar\omega}\left(2\frac{E}{\hbar\omega}+3\right)-20\left(1-\frac{E}{\hbar\omega}\right)\right).\quad(17)$$

The carrier number distributions $N_m(E,\hbar\omega)=3^{m-1}f_m(E,\hbar\omega)$ for the first four generations of carriers are plotted in Figure 10A and B. As one can see, within roughly time $\tau_{e,cool}\sim 3\tau_{ee}$, the distribution changes dramatically and in fact resembles the distribution one would expect if one used the classical Drude model in which absorption light generates many low-energy carriers via "friction"; however, it is important that in the quantum picture, this does not happen instantly, and hot carriers may depart the metal before they decay. Also, even for the fourth generation of carriers, one cannot introduce equilibrium temperature T_e, as evident from the Figure 10B where the negative slope of the distribution increases at higher energies, indicating a reduced number of high-energy carriers capable of surpassing the energy barrier.

Let us now estimate the chances for the hot carriers of each generation to overcome a potential barrier Φ. Two cases will be considered. In the first case, we assume that the transverse momentum is conserved and the efficiency of carrier extraction is [6, 7]

$$\eta^{(+)}_{ext,m}(\Phi,\omega)\sim\frac{m_s}{2m_mE_F}\int_{\Phi}^{\hbar\omega}(E-\Phi)N_m(E,\hbar\omega)\,dE,\quad(18)$$

where m_s and m_m are the effective masses of metal. The results are shown in Figure 11A (without the term in front of the integral, as we are only interested in the relative strength of the injection of carriers from different generations). As one can see, the probability of extraction decreases dramatically in each generation for the barrier height that is at least 30% of the photon energy.

For the second case, we assume that the momentum conservation rules are fully relaxed, and, therefore, all we need is to evaluate the total number of the carriers with energy above the barrier [33]

$$\eta^{(-)}_{ext,m}(\Phi,\omega)\sim C\int_{\Phi}^{\hbar\omega}N_m(E,\hbar\omega)\,dE,\quad(19)$$

where C is the band structure-dependent factor to be derived in the next section; however, at this point, we are interested only in the relative impact of the secondary electrons and holes. The results are shown in Figure 11B and are similar to the case of complete momentum conservation, although the secondary carriers become important for the barriers that are less than half of the photon energy. We summarize the impact of extraction of

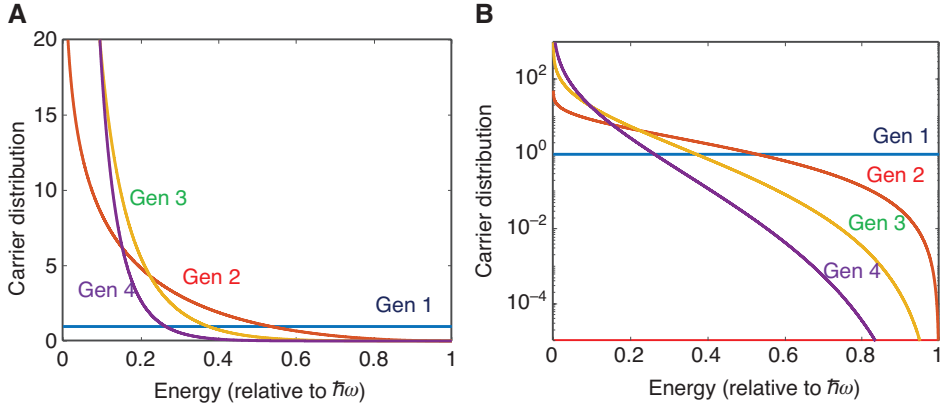

Figure 10: Energy distribution of the first through fourth generation of carriers generated as a result of decay of a single SPP with energy $\hbar\omega$.
(A) linear scale; (B) logarithmic scale. Note that the distribution cannot be defined by a single electron temperature T_e.

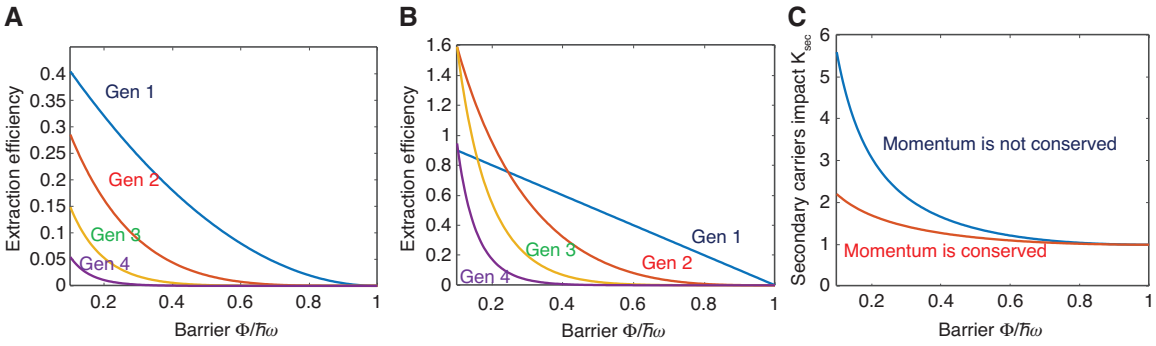

Figure 11: Relative extraction efficiency of the primary and secondary generations of non-equilibrium carriers of different generations as a function of barrier height.
(A) momentum conservation is enforced; (B) momentum conservation is violated; (C) impact of the secondary carriers.

secondary electrons in Figure 11C, where we plot the function $K_{sec} = \sum_{m=1}^{4} \eta_{ext,m} / \eta_{ext,1}$. As one can see, for the barrier that is at least half as large as the phonon energy, the impact of the secondary electrons is negligible, no matter what model we assume.

For the IR detectors, which are one of the more promising hot carrier applications, it is desirable to have the barrier height relatively high to reduce thermal noise, and, as shown in Ref. [77], maximum detectivity is achieved at $\hbar\omega - \Phi \approx 4k_B T << \hbar\omega$. Hence, the impact of secondary carriers can be completely neglected – for all practical purposes, once a single EE scattering event takes place, the carriers are no longer capable of overcoming the barrier. For other cases where the barrier is relatively low, one can simply use the semi-empirical expression to modify the time it takes the primary (first-generation) carriers to decay to the point where they no longer overcome the barrier, as $\tau_{ee}^{eff} = K_{sec}\tau_{ee}$, where K_{sec} is typically <2.

These results are also relevant to the carriers generated via EE-assisted and interband absorption. The holes generated via interband absorption in the d-shell can decay into two holes and one electron, all in the s-band, where they can move relatively fast, and some of those carriers may have energy sufficient to exit across the barrier. The energy distribution of these second-generation carriers is similar to that shown in Figure 9 for $E_1 = E_{ds}$. As long as the barrier is close to the photon energy, one can completely neglect the injection of the intraband-absorption-generated carriers; otherwise, one can simply add their relatively small contribution to K_{sec}. At any rate, once interband absorption commences, the Q of the SPP mode decreases and so does the field enhancement, thus negating the whole goal of plasmonic-assisted detection or catalysis. Similarly, judging from Figure 4, one cannot expect a large contribution from the carriers generated with the help of EE scattering. Once again, that contribution can definitely be ignored for IR light; for visible light, the contribution can also be incorporated into K_{sec}.

Thus, to conclude this section, we state that for all practical purposes, only the primary (first-generation) carriers generated with the phonon/defect assistance or via LD are the ones that can find their way out of the metal. Once these carriers undergo a single EE scattering event, their energies will, for the most part, be way too small to overcome the barrier on the metal/semiconductor (dielectric) interface. We shall refer to these carriers as "quasi-ballistic," as they are expected to propagate quasi-ballistically (phonon and defect scattering does not reduce energy significantly) toward the interface and then get ejected across the barrier. The distribution into which the secondary electrons created as result of EE scattering eventually settle cannot be characterized by a single electron temperature T_e, and, practical values of the incident light intensity never contribute to the injection into the semiconductor over a reasonably high barrier. It is harder to speculate whether this conclusion also holds for the process of photocatalysis on the surface of the metal, as these processes are not yet entirely understood. Still, for a reasonably high activation energy, it seems that only the quasi-ballistic carriers have sufficient energy to initiate the chemical reaction.

4 Exodus. Hot carriers are ejected from metal into the semiconductor or dielectric. How efficiently?

4.1 Transport efficiency

Let us now establish the efficiency of the hot electron injection, $\eta_{ext}(\Phi,\omega) = N_{ext}/N_{SPP}$, where N_{ext} is the number of carriers injected into semiconductor/dielectric. This

efficiency can be split into two factors: the transport efficiency $\eta_{tran}(\omega)=N_s/N_{\rm SPP}$, where N_s is the number of carriers reaching the surface of the metal, and the extraction efficiency $\eta_{ext}(\Phi,\omega)=N_{inj}/N_s$.

To estimate the transport efficiency, we first introduce the mean free path of hot carriers, $L_{mfp,e}=v_F\tau_{ee}^{\rm eff}$. Note that this definition is different from the mean free path distribution in the Drude transport theory, as it involves only EE scattering, because, as mentioned above, the collisions with phonons or defects do not affect the energy of hot carriers. Then, we can introduce the "surface proximity factor":

$$\Gamma_{prox}=\frac{\int\limits_{\rm metal}\mathcal{E}(\mathrm{r})^2\langle\exp(-R/L_{mfp,e})\rangle_\Omega\,\mathrm{d}V}{\int\limits_{\rm metal}\mathcal{E}(\mathrm{r})^2\,\mathrm{d}V}\le 1, \tag{20}$$

where R is the distance to the surface and averaging is done over the solid angle. Obviously, for small nanoparticles with dimension less than L_{mfp}, the proximity factor approaches unity. With that, the transport efficiency becomes

$$\eta_{tran}=\frac{\gamma_{\rm LD}+\Gamma_{prox}\gamma_{ph}}{\gamma_{ib}+\gamma_{ee}+\gamma_{\rm LD}+\gamma_{ph}+\gamma_{rad}}, \tag{21}$$

where γ_{rad} is the radiative decay rate, which is rather small for nanoparticles with radius <50 nm as well as for the propagating SPP mode. We assume that operates below the offset of interband absorption, and, with insignificant EE scattering, one can obtain the expression for the spherical nanoparticle.

For the case of propagating SPP in which the intensity inside the metal decays exponentially as $\exp(-x/L_p)$, one can estimate $\Gamma_{prox}=L_{mfp,e}/(L_p+L_{mfp,e})$, where L_p is the penetration depth and one obtains

$$\eta_{tran}=\frac{1}{2}\frac{\frac{3}{8}\frac{v_F}{L_p}+\frac{L_{mfp,e}}{L_p+L_{mfp,e}}\frac{1}{\tau_{ep}}}{\frac{3}{8}\frac{v_F}{L_p}+\frac{1}{\tau_{ep}}}=\frac{1}{2}-\frac{L_p^2}{2(L_p+L_{mfp,e})(L_p+\frac{3}{8}L_{mfp,p})}, \tag{22}$$

where the mean free path due to phonon and defect collisions is $L_{mfp,p}=v_F\tau_{ep}$, and the factor 1/2 accounts for the fact that only one-half of the hot carriers move toward the surface. For the Au or Ag guide in the near-IR range where the Drude approximation for the dielectric constant $\varepsilon_m(\lambda)\approx 1-\lambda^2/\lambda_p^2$, the penetration depth $L_p\approx\lambda_p/4\pi\approx 12$ nm, where $\lambda_p=140$ nm is the plasma wavelength. For $\tau_{ee}^{\rm eff}=10$ fs and $\tau_{ep}=15$ fs, we obtain $L_{mfp,e}=14$ nm, $L_{mfp,p}=40$ nm, and $\eta_{tran}\approx 40\%$. For the case of a small spherical nanoparticle with diameter d, we obtain $\Gamma_{prox}\approx 0.7\exp(-d/2L_{mfp,e})$ and

$$\eta_{tran}=\frac{\frac{3}{8}\frac{v_F}{d}+0.7\exp(-d/2L_{mfp,e})\frac{1}{\tau_{ep}}}{\frac{3}{8}\frac{v_F}{d}+\frac{1}{\tau_{ep}}}=\frac{1+1.84\exp(-d/2L_{mfp,e})d/L_{mfp,p}}{1+2.66d/L_{mfp,p}}, \tag{23}$$

where we have neglected the small possibility of the LD carriers generated at one end of a nanoparticle going all the way to the other end without scattering. For a $d=20$ nm nanosphere, one gets $\eta_{tran}\approx 52\%$ and for $d=40$ nm, $\eta_{tran}\approx 30\%$. As carriers generated by EE and phonons have different angular distributions, the overall distribution of carriers near the surface (shown in Figure 5D) is

$$R_{eff}(\theta)\approx a_{ph}R_{ph}(\theta)+(1-a_{ph})R_{LD}(\theta)=a_{ph}\left(\frac{3}{4}\cos^2\theta+\frac{1}{4}\right)+2(1-a_{ph})|\cos^3\theta|, \tag{24}$$

where a_{ph} is the fraction of carriers generated via phonon/defect scattering and according to Eq. (22) for the propagating SPP $a_{ph}\approx 0.45$, while according to Eq. (24) for spherical nanoparticles with 12 nm < d < 60 nm a_{ph} stays between 0.4 and 0.45. In other words, phonon- and defect-assisted absorption is responsible for almost one-half of the ballistic carriers arriving at the metal surface, and one can write the effective angular distribution as $R_{eff}(\theta)\approx 1.1|\cos^3\theta|+0.33\cos^2\theta+0.11$, as shown in Figure 5D.

4.2 Extraction efficiency

With transport to the surface out of the way, we consider the extraction efficiency of all the quasi-ballistic carriers arriving at the surface, i.e. the transmission coefficient over the barrier Φ. As shown in Figure 12A, if the lateral (in plane) wavevector is continuous across the barrier, i.e. $\boldsymbol{k}_{m,\|}=\boldsymbol{k}_{s,\|}=\boldsymbol{k}_{\|}$, then the longitudinal wavevector for the electron in the metal whose energy above the Fermi level is E and whose incidence angle is θ can be found as $k_{m,z}=\sqrt{2m_m/\hbar^2(E+E_F)-k_{\|}^2}\approx k_F\cos\theta$, where m_m is the effective mass of metal. For the semiconductor, $k_{s,z}\approx k_F\sqrt{\sin^2\theta_{\max}-\sin^2\theta}$, where $\theta_{\max}(E,\Phi)=\sin^{-1}\sqrt{(m_s/m_m)(E-\Phi)/E_F}$ and m_s is the effective mass of semiconductor. As a result of wavevector continuity

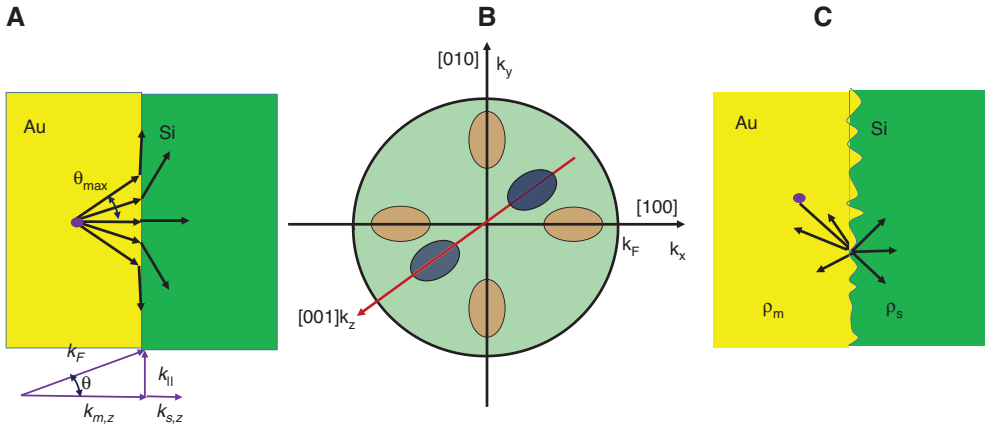

Figure 12: Impact of interface rougness on carrier injection.
(A) Carrier injection from the metal into silicon across a smooth interface. Only electrons with an incidence angle less than θ_{max} get injected. (B) Six valleys in the conduction band of Si. Electrons get injected only into the two valleys along [100] direction. (C) Carrier injection from the metal into silicon across a rough interface with momentum conservation restriction entirely lifted.

(lateral momentum conservation), only the carriers with $\theta \le \theta_{max}(E,\Phi)$ can be extracted from the metal. First, we assume that all carriers within the "extraction cone" can exit the metal; hence

$$\eta_{ext}(\hbar\omega,\Phi) = \int_0^{\hbar\omega} f_1(E,\hbar\omega) \int_0^{\theta_{max}(E,\Phi)} R_{eff}(\theta)\sin\theta\, d\theta\, dE, \qquad (25)$$

where $f_1(E, \hbar\omega) = 1/\hbar\omega$ according to Eq. (17). In our calculations, we consider injection from a noble metal, like Au into Si. The conduction band of Si is characterized by six valleys along <100> directions, each characterized by longitudinal $m_L = 0.98m_0$ and transverse $m_T = 0.19m_0$, as shown in Figure 12B. The barrier is treated as a variable parameter due to the presence of surface states. The effective injection is possible almost exclusively in two valleys along [100] (normal to the interface) directions, as, in the other four valleys, the Bloch functions have a symmetry that is almost orthogonal to the S-states in the metal. We then approximate the Si conduction band with a single valley isotropic band with an effective DOS mass of $m_s = 2^{2/3} m_L^{1/2} m_T^{2/3} = 0.52m_0$. If the injection is into the valence band, the effective mass is almost exactly the same, $m_s = 0.49m_0$.

Then, as typically θ_{max} is rather small, we can neglect the angular dependence of the ballistic carrier distribution, assuming $R_{eff}(\theta) \approx R_{eff}(0)$ and obtain

$$\eta_{ext}(\hbar\omega, \Phi) \approx \frac{1}{4} R_{eff}(0) \frac{m_s}{m_0} \frac{(\hbar\omega - \Phi)^2}{\hbar\omega E_F}. \qquad (26)$$

or essentially a Fowler's formula [78], as plotted in Figure 13A for $\hbar\omega = 0.8$ eV ($\lambda = 1500$ nm) and is practically no different from the exact formula (25).

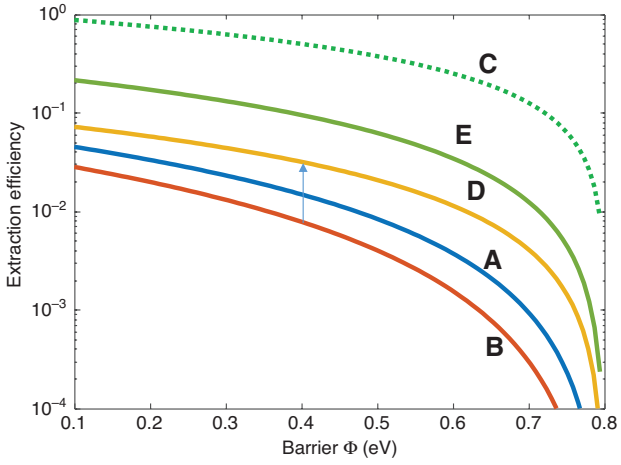

Figure 13: Extraction efficiency of the hot carriers generated on the metal/Si interface by the SPPs excited with $\lambda = 1500$ nm photons. (A) Smooth interface; reflection is not taken into account [Eq. (27)]. (B) Smooth interface; reflection is taken into account [Eq. (27)]. (C) Rough interface; complete extraction of all above-the-barrier carriers [Eq. (28)]. (D) Rough interface; momentum conservation rules are relaxed, but transitions into transverse valleys of Si are not allowed [Eq. (29)]. (E) Rough interface; momentum conservation rules are relaxed and the transitions into transverse valleys of Si are allowed [Eq. (29)].

Next, we shall take into consideration the reflection from the interface and include transmission coefficient

$$\eta_{ext}(\hbar\omega, \Phi) = \int_0^{\hbar\omega} f_1(E, \hbar\omega) \int_0^{\theta_{max}(E,\Phi)} \left[1 - \left(\frac{k_{m,z}/m_0 - k_{s,z}/m_s}{k_{m,z}/m_0 + k_{s,z}/m_s}\right)^2\right] R_{eff}(\theta)\sin\theta\, d\theta\, dE \qquad (27)$$

into the integral in Eq. (25) – the result is shown in Figure 13B, and, as one can see, the extraction efficiency

is reduced by about 20% for low barriers but by as much as a factor of 2 for a barrier that is only 200 meV below the photon frequency.

As one can see, lateral momentum conservation severely restricts the "exit cone" of the incident ballistic carriers to the small angle $\theta_{\max} < \pi/10$ for the visible and even less for near-IR wavelengths leading to small extraction efficiencies. However, experimental data show that higher injection efficiencies can be achieved when the momentum conservation is no longer valid due to extreme disorder at the interface. In Ref. [33], an injection efficiency of nearly 30% for Au/GaAs interface was reported, while high efficiencies for injection into TiO_2 from Au nanoparticles have been measured in Ref. [34]. Increase in photocurrent in the photodetectors with a rough Au/Si interface relative to the ones with a smooth interface has been reported in Ref. [79].

To explain these extraordinary results, the simplest model was proposed in Ref. [33] that assumed that all the hot carriers with energies higher than barrier Φ can be extracted as in Eq. (19), leading to

$$\eta_{ext,\max}(\hbar\omega, \Phi) = \frac{\hbar\omega - \Phi}{\Phi}, \tag{28}$$

as shown in Figure 13C; however, this approach entirely neglects the possibility of backscattering into the metal. Another approach [79] is to use the explicit interface roughness scattering explicitly to obtain the enhancement of extraction efficiency by a factor of a few. That model, however, could only be applied to a relatively small roughness, and, as matter of fact, neglects enhanced backscattering as well.

To find the ultimate extraction efficiency, we shall follow the theory of Yablonovitch [45] developed for the seemingly different task of light trapping in a dielectric with a roughened surface. Essentially, the argument developed there can be applied to the case of the surface roughened to the degree that momentum conservation is no longer valid – then, according to Fermi golden rule, the rate of scattering in a given direction depends only on the DOS, as shown in Figure 12C. Now, if the densities of states in the metal and semiconductor are ρ_m and ρ_s, respectively, one can obtain a rather simple expression for the extraction efficiency:

$$\eta_{ext,\max}(\hbar\omega, \Phi) = \int_0^{\hbar\omega} f_1(E, \hbar\omega) \frac{\rho_s(E)}{\rho_s(E) + \rho_m(E)} dE = \frac{1}{\hbar\omega} \int_0^{\hbar\omega} \frac{(m_s/m_0)^{3/2}(E-\Phi)^{1/2}/E_F^{1/2}}{(m_s/m_0)^{3/2}(E-\Phi)^{1/2}/E_F^{1/2} + 1} dE. \tag{29}$$

For small extraction probability, one can obtain the estimate

$$\eta_{ext}(\hbar\omega, \Phi) = \frac{2}{3}\left(\frac{m_s}{m_0}\right)^{3/2} \frac{(\hbar\omega - \Phi)^{3/2}}{\hbar\omega E_F^{1/2}}. \tag{30}$$

As one can see, the result [Eq. (30)] is not nearly as high as the estimate [Eq. (28)], yet it is higher than the extraction efficiency without roughness [Eq. (26)] by roughly a factor. For Si, it is important to consider what value of effective DOS mass to use in Eqs. (29) and (30). If one assumes that the electron can be injected only into two X-valleys, one should use the previously defined effective mass, $m_s = 0.52m_0$. However, it is possible for the disorder to be so strong that it allows transitions into the other four valleys, which would imply a DOS mass equal to $m_s = 1.08m_0$, increasing the extraction efficiency by a factor of 3. The two curves are plotted in Figure 13D and E, respectively. Whether one can induce such a strong disorder that it will break the selection rules that prevent the injection into the "transverse valleys" is difficult to state, but it is definitely not inconceivable. Therefore, if one assumes that the barrier height is 0.4 eV, one can see that introducing roughness can increase the extraction efficiency by a factor between 4 and 12, from <1% to >10%, leading to an overall efficiency of 4–5%. The increase is even more dramatic for higher barriers, which is where the performance of the detector becomes optimal due to a decrease of dark current.

In Figure 14, we show the extraction efficiency with and without disorder for the case of an Au/TiO_2 interface and a wavelength of 620 nm (photon energy of 2 eV). The effective mass of TiO_2 is $m_s = 0.8m_0$ and, for a relatively high barrier, one gets significant improvement of the extraction efficiency due to scattering on a rough interface. However, typically, the barrier height on the metal-TiO_2 interface is only a few hundred meV and roughness only increases η_{ext} by a factor of 2 or so. Overall, one can expect extraction efficiencies on the scale of 10–20%. Thus, the fact that with TiO_2 one attains a higher injection efficiency can be traced to the fact that it has a large DOS available for injection.

Overall, the injection efficiency for Au/TiO_2 can be as high as 10%, or even higher, if one assumes that the carriers can travel from one side of nanoparticle to the other. For Au/Si, the number is smaller, primarily due to smaller DOS and higher barrier, and is typically on the scale of a few percent. Thus, while the quantum efficiency of hot carrier detectors can never reach those easily attainable in commercial photodiodes, it may be sufficient for these

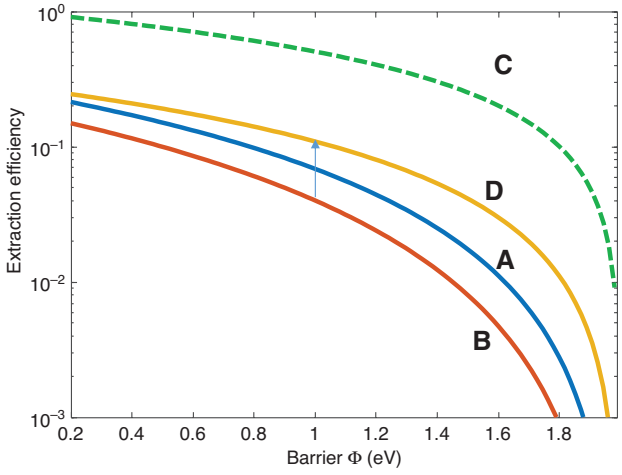

Figure 14: Extraction efficiency of the hot carriers generated on the metal/TiO_2 interface by the SPPs excited with $\lambda = 620$ nm photons. (A) Smooth Interface; reflection is not taken into account [Eq. (26)]. (B) Smooth Interface; reflection is taken into account [Eq. (27)]. (C) Rough interface; complete extraction of all above-the-barrier carriers [Eq. (28)]. (D) Rough interface; momentum conservation rules are relaxed [Eq. (30)].

detectors to find some applications because they can absorb long-wave radiation while being compatible with commercial silicon technology. Whether the same can be said about plasmon-assisted photocatalysis remains to be seen.

5 Conclusions

At this point, it is time to summarize the main points made in the present work so the busy reader can be spared the arduous task of going through all the derivations and calculations performed in prior sections and instead focus on the practical consequences of this work.

The first important point made in this work relates to the inherently discrete, quantum nature of SPP generation and decay. Under realistic illumination, with optical power densities far less than 1 MW/cm^2 at any given time, no more than a single SPP is present on a typical nanoparticle (the same is true for a number of SPPs inside a plasmonic photodetector waveguide when the optical power is <10 μW). Consequently, at any given time, the number of non-equilibrium carriers in a given nanoparticle (or inside a waveguide) is typically only a few and their combined energy is exactly $\hbar\omega$. Therefore, one cannot use the "average" electron temperature T_e to describe the evolution of hot carriers, as their distribution is never thermal. The one and only way to describe the evolution of hot carriers is to simply follow them through the scattering events, generation after generation.

The second point is that as long as the height of barrier separating the metal from the adjacent isolator/semiconductor is at least moderately large (higher than $\hbar\omega/3$), only two SPP decay mechanisms out of four, namely LD and phonon-/defect-assisted decay, generate a significant number of carriers with energies high enough to surpass the barriers. Furthermore, after only a single event of EE scattering, a sufficiently energetic first-generation carrier decays into three carriers whose energies are too low for the extraction across the barrier. Thus, hot carriers have only a very short (~10 fs) time over which they can be injected from the metal into the semiconductor/isolator. After that, no injection can take place. For this reason, the carriers generated at the surface via LD stand the highest chance of being injected.

The third point is that the extraction efficiency of hot carriers is greatly affected by the smoothness of the interface. For a smooth interface, momentum conservation dictates that only a very small fraction of hot carriers with small in-plane wavevectors are capable of exiting the metal. Typical injection efficiencies do not exceed 0.1%. However, if the surface roughness is so high that momentum is no longer conserved, practically all the carriers with energies above the barrier stand a chance of exiting this metal. However, this chance is far from 100% because the DOS near Fermi in the metal is typically much higher than in the semiconductor. Still, depending on the barrier height and effective mass, up to 10% injection efficiencies are possible. Whether the momentum is conserved or not, it is desirable to have a semiconductor/dielectric with large effective mass (and thus DOS) and a metal with lower DOS near Fermi level, which may make alternative plasmonic materials [80], such as TiN, attractive.

Finally, when it comes to photodetectors, the main practical parameter is detectivity, and, for that, one must reduce thermal noise; hence, one does not benefit from lowering the barrier beyond the optimal 4 kT. Choosing a semiconductor with large DOS, such as Si or Ge, rather than, say, GaAs, remains to be the only viable stratagem for the performance improvement in addition to the aforementioned roughening of the interface and engineering the waveguide mode to make sure that the field is concentrated near the surface and that LD is a dominant SPP decay channel. For the photocatalysis, one should also increase the relative strength of LD by using smaller nanoparticles with large surface-to-volume ratio and roughen the interface, and, as thermal noise is not a factor, the barrier should be lowered. However, when the barrier is sufficiently low, most of the enhancement of catalysis

will come not from hot carriers per se but simply from the thermionic emission due to the increase of the ambient temperature, and heating can be achieved by means other than light absorption.

In the end, I have presented here a compilation of factors determining the efficiency of plasmonic-assisted hot carrier injection for applications in (mostly) detectors and (also) photocatalysis. Some of the results presented here have been, of course, investigated before. For instance, the fact that non-equilibrium carrier distribution is not thermal has been argued by many. Nevertheless, other results are indeed entirely new. In particular, the important fact that for realistic illumination conditions no more than a single SPP gets excited per nanoparticle, surprisingly, has been overlooked before. Also, the paramount role of DOS in the injection process has not been given proper attention. It is my belief that it is valuable for the plasmonic community to combine in one place a coherent and unified description of all the steps of the process – from SPP generation through their decay engendering non-equilibrium carriers that then go through competing processes of decay, transport to the surface, and extraction from metal. This treatment provides a simple way of estimating the overall efficiency of the injection and also outlines the pathways to its optimization. Whether the research community finds any value in this modest effort is an open question, but I sure hope that not far in the future hot carrier devices will enter the mainstream and this work will play a helpful role, no matter how small, in it.

Acknowledgments: The support of a National Science Foundation grant (no. 1507749, Funder Id: http://dx.doi.org/10.13039/100000001) has been a key to writing this article. Aside from that, valuable if sometimes heated discussions with Prof. P. Noir as well as finer points made by the new member of his research group, Ms. S. Artois, have been essential to this work.

References

[1] Stockman MI. Nanoplasmonics: past, present, and glimpse into future. Opt Express 2011;19:22029–106.
[2] Gramotnev DK, Bozhevolnyi SI. Plasmonics beyond the diffraction limit. Nat Photonics 2010;4:83–91.
[3] Atwater HA, Polman A. Plasmonics for improved photovoltaic devices. Nat Mater 2010;9:205–13.
[4] Maier SA. Plasmonics: fundamentals and applications. New York: Springer, 2007: xxiv, 223 p.
[5] Kim JB, Lee J-H, Moon C-K, Kim S-Y, Kim J-J. Highly enhanced light extraction from surface plasmonic loss minimized organic light-emitting diodes. Adv Mater 2013;25:3571–7.
[6] Goykhman I, Desiatov B, Khurgin J, Shappir J, Levy U. Locally oxidized silicon surface-plasmon Schottky detector for telecom regime. Nano Lett 2011;11:2219–24.
[7] Goykhman I, Desiatov B, Khurgin J, Shappir J, Levy U. Waveguide based compact silicon Schottky photodetector with enhanced responsivity in the telecom spectral band. Opt Express 2012;20:28594–602.
[8] Haffner C, Chelladurai D, Fedoryshyn Y, et al. Low-loss plasmon-assisted electro-optic modulator. Nature 2018;556:483–6.
[9] Melikyan A, Vallaitis T, Lindenmann N, Schimmel T, Freude W, Leuthold J. Surface plasmon polariton absorption modulator. Opt Express 2011;19:8855–69.
[10] Zhang SP, Bao K, Halas NJ, Xu H, Nordlander P. Substrate-induced Fano resonances of a plasmonic nanocube: a route to increased-sensitivity localized surface plasmon resonance sensors revealed. Nano Lett 2011;11:1657–63.
[11] Stewart ME, Anderton CR, Thompson LB. Nanostructured plasmonic sensors. Chem Rev 2008;108:494–521.
[12] Khurgin JB. How to deal with the loss in plasmonics and metamaterials. Nat Nanotechnol 2015;10:2–6.
[13] Khurgin JB. Replacing noble metals with alternative materials in plasmonics and metamaterials: how good an idea? Philos Trans A Math Phys Eng Sci 2017;375:20160068.
[14] Hsieh WT, Wu PC, Khurgin JB, Tsai DP, Liu N, Sun G. Comparative analysis of metals and alternative infrared plasmonic materials. ACS Photonics 2018;5:2541–8.
[15] Citroni R, Di Paolo F, Di Carlo A. Replacing noble metals with alternative metals in MID-IR frequency: a theoretical approach. Nanoinnovation 2017;2018:1990.
[16] Moskovits M. The case for plasmon-derived hot carrier devices. Nat Nanotechnol 2015;10:6–8.
[17] Brongersma ML, Halas NJ, Nordlander P. Plasmon-induced hot carrier science and technology. Nat Nanotechnol 2015;10:25–34.
[18] Linic S, Christopher P, Ingram DB. Plasmonic-metal nanostructures for efficient conversion of solar to chemical energy. Nat Mater 2011;10:911–21.
[19] Knight MW, Sobhani H, Nordlander P, Halas NJ. Photodetection with active optical antennas. Science 2011;332:702–4.
[20] Panchenko E, Cadusch JJ, James TD, Roberts A. Plasmonic metasurface-enabled differential photodetectors for broadband optical polarization characterization. ACS Photonics 2016;3:1833–9.
[21] Lal S, Clare SE, Halas NJ. Nanoshell-enabled photothermal cancer therapy: impending clinical impact. Acc Chem Res 2008;41:1842–51.
[22] Zhou Z, Sakr E, Sun Y, Bermel P. Solar thermophotovoltaics: reshaping the solar spectrum. Nanophotonics 2016;5:1–21.
[23] Christopher P, Xin HL, Linic S. Visible-light-enhanced catalytic oxidation reactions on plasmonic silver nanostructures. Nat Chem 2011;3:467–72.
[24] Zhang C, Zhao HQ, Zhou LA, et al. Al-Pd nanodisk heterodimers as antenna-reactor photocatalysts. Nano Lett 2016;16:6677–82.
[25] Mukherjee S, Libisch F, Large N, et al. Hot electrons do the impossible: plasmon-induced dissociation of H-2 on Au. Nano Lett 2013;13:240–7.
[26] Zhang YC, He S, Guo WX, et al. Surface-plasmon-driven hot electron photochemistry. Chem Rev 2018;118:2927–54.
[27] Clavero C. Plasmon-induced hot-electron generation at nanoparticle/metal-oxide interfaces for photovoltaic and photocatalytic devices. Nat Photonics 2014;8:95–103.

[28] Manjavacas A, Liu JG, Kulkarni V, Nordlander P. Plasmon-induced hot carriers in metallic nanoparticles. ACS Nano 2014;8:7630–8.
[29] Rethfeld B, Kaiser A, Vicanek M, Simon G. Ultrafast dynamics of nonequilibrium electrons in metals under femtosecond laser irradiation. Phys Rev B 2002;65:214303.
[30] Sundararaman R, Narang P, Jermyn AS, Goddard III WA, Atwater HA. Theoretical predictions for hot-carrier generation from surface plasmon decay. Nat Commun 2014;5:5788.
[31] Brown AM, Sundararaman R, Narang P, Goddard 3rd WA, Atwater HA. Nonradiative plasmon decay and hot carrier dynamics: effects of phonons, surfaces, and geometry. ACS Nano 2016;10:957–66.
[32] Hartland GV, Besteiro LV, Johns P, Govorov AO. What's so hot about electrons in metal nanoparticles? ACS Energy Lett 2017;2:1641–53.
[33] Giugni A, Torre B, Toma A, et al. Hot-electron nanoscopy using adiabatic compression of surface plasmons. Nat Nanotechnol 2013;8:845–52.
[34] Ratchford DC, Dunkelberger AD, Vurgaftman I, Owrutsky JC, Pehrsson PE. Quantification of efficient plasmonic hot-electron injection in gold nanoparticle TiO_2 films. Nano Lett 2017;17:6047–55.
[35] Zhou LA, Swearer DF, Zhang C, et al. Quantifying hot carrier and thermal contributions in plasmonic photocatalysis. Science 2018;362:69–72.
[36] Govorov AO, Zhang H, Gun'ko YK. Theory of photoinjection of hot plasmonic carriers from metal nanostructures into semiconductors and surface molecules. J Phys Chem C 2013;117:16616–31.
[37] Besteiro LV, Kong X-T, Wang Z, Hartland GV, Govorov AO. Understanding hot-electron generation and plasmon relaxation in metal nanocrystals: quantum and classical mechanisms. ACS Photonics 2017;4:2759–81.
[38] Ng C, Cadusch JJ, Dligatch S, et al. Hot carrier extraction with plasmonic broadband absorbers. ACS Nano 2016;10:4704–11.
[39] Protsenko IE, Uskov AV. Photoemission from metal nanoparticles. Physics-Uspekhi 2012;55:508–18.
[40] Sivan Y, Baraban J, Un IW, Dubi Y. Comment on "Quantifying hot carrier and thermal contributions in plasmonic photocatalysis". Science 2019;364.
[41] Sivan Y, Un IW, Dubi Y. Assistance of metal nanoparticles in photocatalysis – nothing more than a classical heat source. Faraday Discuss 2019;214:215–33.
[42] Narang P, Sundararaman R, Atwater HA. Plasmonic hot carrier dynamics in solid-state and chemical systems for energy conversion. Nanophotonics 2016;5:96–111.
[43] Christopher P, Moskovits M. Hot charge carrier transmission from plasmonic nanostructures. Annu Rev Phys Chem 2017;68:379–98.
[44] Khurgin JB. Hot carriers generated by plasmons: where are they generated and where do they go from there? Faraday Discuss 2019;214:35–58.
[45] Yablonovitch E. Statistical ray optics. J Opt Soc Am 1982;72:899–907.
[46] Deckman HW, Roxlo CB, Yablonovitch E. Maximum statistical increase of optical-absorption in textured semiconductor-films. Optics Lett 1983;8:491–3.
[47] Khurgin JB, Sun G, Soref RA. Practical limits of absorption enhancement near metal nanoparticles. Appl Phys Lett 2009;94:071103.
[48] Li K, Stockman MI, Bergman DJ. Self-similar chain of metal nanospheres as an efficient nanolens. Phys Rev Lett 2003;91:227402.
[49] Nordlander P, Oubre C, Prodan E, Li K, Stockman MI. Plasmon hybridization in nanoparticle dimers. Nano Lett 2004;4:899–903.
[50] Petek H, Ogawa S. Femtosecond time-resolved two-photon photoemission studies of electron dynamics in metals. Progr Surf Sci 1997;56:239–310.
[51] Schmuttenmaer CA, Aeschlimann M, Herman JW, et al. Femtosecond studies of carrier relaxation processes at single-crystal metal-surfaces. In: Proc. SPIE 2125, Laser Techniques for Surface Science 1994;2125:98–106.
[52] Lugovskoy AV, Bray I. Ultrafast electron dynamics in metals under laser irradiation. Phys Rev B 1999;60:3279–88.
[53] Ono S. Thermalization in simple metals: role of electron-phonon and phonon-phonon scattering. Phys Rev B 2018;97:054310.
[54] Ziman JM. Electrons and phonons: the theory of transport phenomena in solids. Oxford classic texts in the physical sciences. Oxford, New York: Clarendon Press; Oxford University Press, 2001: xiv, 554 p.
[55] Bouillard JSG, Dickson W, O'Connor DP, Wurtz GA, Zayats AV. Low-temperature plasmonics of metallic nanostructures. Nano Lett 2012;12:1561–5.
[56] Jayanti SV, Park JH, Dejneka A, et al. Low-temperature enhancement of plasmonic performance in silver films. Opt Mater Express 2015;5:1147–55.
[57] Theye ML. Investigation of the optical properties of Au by means of thin semitransparent films. Phys Rev B Solid State 1970;2:3060–78.
[58] Abelès F. Optical properties of solids. Amsterdam, New York: North-Holland Pub. Co., American Elsevier, 1972: 1034 p.
[59] Lawrence WE, Wilkins JW. Electron-electron scattering in transport coefficients of simple metals. Phys Rev B 1973;7:2317–32.
[60] Pines D, Nozières P. The theory of quantum liquids. New York: W.A. Benjamin, 1966: v.
[61] Beach RT, Christy RW. Electron-electron scattering in intraband optical conductivity of Cu, Ag, and Au. Phys Rev B 1977;16:5277–84.
[62] Parkins GR, Lawrence WE, Christy RW. Intraband optical conductivity sigma(omega,tau) of Cu, Ag, and Au – contribution from electron-electron scattering. Phys Rev B 1981;23:6408–16.
[63] Uskov AV, Protsenko IE, Ikhsanov RS, et al. Internal photoemission from plasmonic nanoparticles: comparison between surface and volume photoelectric effects. Nanoscale 2014;6:4716–27.
[64] Uskov AV, Protsenko IE, Asger Mortensen N, O'Reilly EP. Broadening of plasmonic resonance due to electron collisions with nanoparticle boundary: a quantum mechanical consideration. Plasmonics 2014;9:185–92.
[65] Khurgin JB. Ultimate limit of field confinement by surface plasmon polaritons. Faraday Discuss 2015;178:109–22.
[66] Khurgin J, Tsai W-T, Tsai DP, Sun G. Landau damping and limit to field confinement and enhancement in plasmonic dimers. ACS Photonics 2017;4:2871–80.
[67] Yannouleas C, Broglia RA. Landau damping and wall dissipation in large metal-clusters. Ann Phys 1992;217:105–41.
[68] Yuan Z, Gao SW. Landau damping and lifetime oscillation of surface plasmons in metallic thin films studied in a jellium slab model. Surf Sci 2008;602:460–4.

[69] Kreibig U, Vollmer M. Optical properties of metal clusters. In: Springer series in materials science. Berlin, New York: Springer, 1995: xx, 532 p.
[70] Shahbazyan TV. Landau damping of surface plasmons in metal nanostructures. Phys Rev B 2016;94:235431.
[71] Shahbazyan TV. Surface-assisted carrier excitation in plasmonic nanostructures. Plasmonics 2018;13:757–61.
[72] Lindhard J. On the properties of a gas of charged particles. Matematisk-Fysiske Meddelelser Kongelige Danske Videnskabernes Selskab 1954;28:1–57.
[73] Fann WS, Storz R, Tom HWK, Bokor J. Electron thermalization in gold. Phys Rev B 1992;46:13592–5.
[74] Fann WS, Storz R, Tom HWK, Bokor J. Direct measurement of nonequilibrium electron-energy distributions in subpicosecond laser-heated gold-films. Phys Rev Lett 1992;68:2834–7.
[75] Bauer M, Marienfeld A, Aeschlimann M. Hot electron lifetimes in metals probed by time-resolved two-photon photoemission. Progr Surf Sci 2015;90:319–76.
[76] Dubi Y, Sivan Y. "Hot" electrons in metallic nanostructures-non-thermal carriers or heating? Light-Sci Appl 2019;8:89.
[77] Grajower M, Desiatov B, Mazurski N, Shappir J, Khurgin JB, Levy U. Optimization and experimental demonstration of plasmonic enhanced internal photoemission silicon Schottky detectors in the mid-IR. ACS Photonics 2017;4:1015–20.
[78] Fowler RH. The analysis of photoelectric sensitivity curves for clean metals at various temperatures. Phys Rev 1931;38:45–56.
[79] Grajower M, Levy U, Khurgin JB. The role of surface roughness in plasmonic-assisted internal photoemission Schottky photodetectors. ACS Photonics 2018;5:4030–6.
[80] Naik GV, Shalaev VM, Boltasseva A. Alternative plasmonic materials: beyond gold and silver. Adv Mater 2013;25:3264–94.