Light at the End of the Tunnel

Originally published in:
ACS Photonics 5(11), https://doi.org/10.1021/acsphotonics.8b00726

Funding acknowledgement:
165841 - Nonlinear optics with hybrid plasmonic-TMDC materials (SNF)

Light at the end of the tunnel

Markus Parzefall* and Lukas Novotny*

Photonics Laboratory, ETH Zürich, 8093 Zürich, Switzerland

E-mail: mparzefall@ethz.ch; lnovotny@ethz.ch

Abstract

In recent years, tunnel junctions have reemerged as promising candidates for the transduction of electrical to optical signals at the nanoscale. The process of interest is known as inelastic electron tunneling (IET), where a tunneling electron excites an optical mode while traversing the tunnel barrier. The main appeal of tunnel junctions lies in their size and bandwidth, both of which are unmatched by other electronic devices. However, their main disadvantage so far has been the overall low transduction efficiency. Recently, the realization of photon [1] and surface plasmon polariton (SPP) [2] sources based on IET—with efficiencies exceeding 1 %—were reported. In this perspective, we critically analyze the factors that limit efficiencies and extract guidelines for the development of efficient photon and SPP sources.

Keywords

inelastic electron tunneling, metal-insulator-metal junctions, surface plasmon polaritons, local density of optical states, optical antennas, optoelectronics

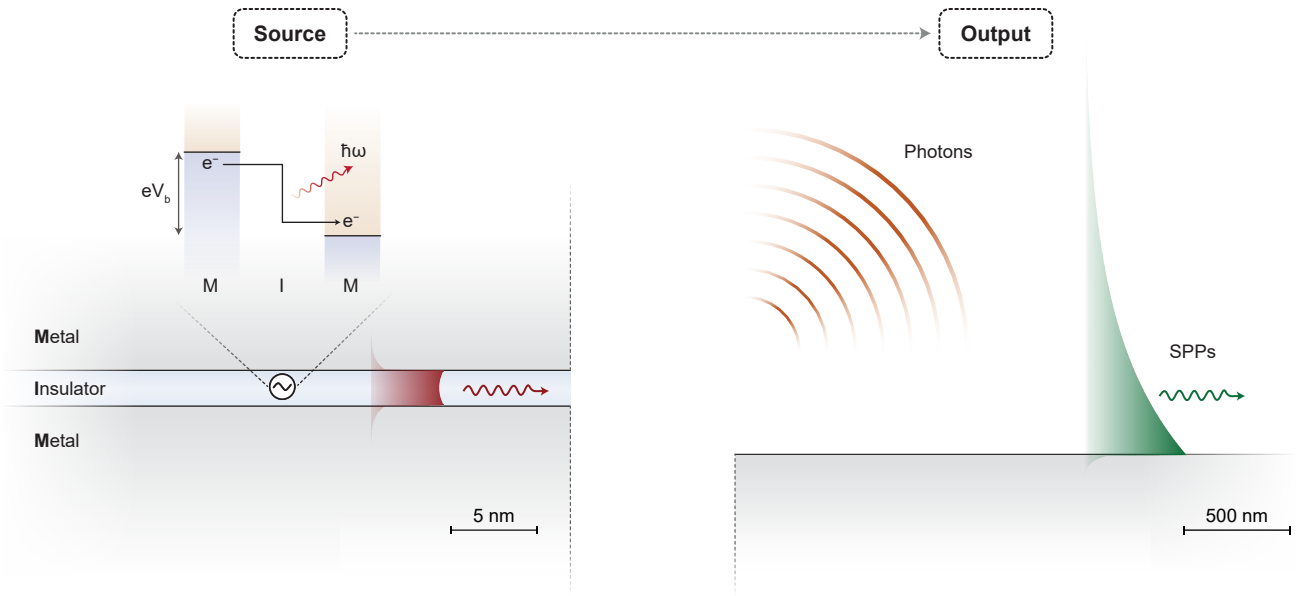

Figure 1: Inelastic electron tunneling in a MIM device. Tunneling electrons act as the source for the excitation of MIM modes, represented by their electric field distribution. We distinguish between two outputs of the system, SPPs or photons. The inset shows a schematic of the IET process.

The excitation of optical modes by quantum tunneling—more specifically inelastic electron tunneling (IET)—has been investigated for four decades. After the initial discovery by Lambe & McCarthy in 1976,[3] IET was studied in (laterally) macroscopic metal-insulator-metal (MIM) devices.[4] Towards the end of the 1980s, IET was rediscovered in the scanning tunneling microscopy (STM) community with the observation of light emission from an STM junction[5,6] and evolved into a spectroscopic tool.[7–9] Most recently, with the advancement of the optical antenna concept,[10,11] renewed interest in IET as a means to electrically excite propagating and localized surface plasmon polariton (SPP) modes emerged.[12,13] In contrast to radio- and microwave antennas, optical antennas are generally excited by light. The coupling of tunnel junctions to optical antennas has eventually allowed for the realization of optical antennas that are driven electrically,[14,15] igniting renewed interest in IET.[16–24]

The common denominator of all these studies is their ultimate focus on the light that is emitted from the (MIM) tunnel junction. While many arguments speak in favor of IET as a driving element for nanoscale light sources, e.g. its scalability and bandwidth, the main

challenge from a practical point of view has always been its low efficiency, i.e. the ratio between the electron tunneling rate and photon emission rate, generally remaining lower than 10^{-3} photons per tunneling electron. Recently, the realization of an IET-driven SPP source with SPP excitation efficiencies exceeding 1 % was reported by Du et al.[2] According to the authors, such comparatively high efficiencies were reached because their devices are not limited by low photon outcoupling efficiencies. Nonetheless, Qian et al. now realized a photon source based on IET, reaching similar values in efficiency.[1]

The scope of this perspective is summarized in Fig. 1. The left hand side shows a vertical cross section of a MIM tunnel junction. The insulating region separating the two metallic domains—typically 1–3 nm thick—serves as the tunnel barrier. The inset shows a schematic of the IET process. A bias V_b supplies electrons in one electrode with excess energy eV_b, with e being the elementary charge. During the tunneling process, an electron may loose some of that excess energy to the excitation of an optical mode with energy $\hbar\omega$, where $\hbar$ is the reduced Planck constant and ω is the angular frequency of the optical mode. The right hand side shows the two potential outputs of the system, photons and SPPs.

In the following we will first derive the source spectrum of IET, i.e. its capacity as a driving source for the optical modes of an arbitrary geometry. As the most widely studied system, we will then discuss the optical properties of the MIM geometry and show that IET acts as a (efficient) drive for the MIM mode. We will further compare the challenges in realizing an efficient photon or SPP source—the two potential output channels. Finally we will lay out guidelines how to engineer the system such that the conversion efficiency between input and output is maximized and discuss recent experimental findings in light of these guidelines.

The source spectrum of IET

The IET process is typically represented as depicted in the inset of Fig. 1, omitting a more specific description of the underlying physical process. Theoretically, IET has been modeled as the energy loss of a quantum mechanical current source,[25–27] as quantum shot noise that originates from fluctuations in the tunneling current[28–31] and as a spontaneous emission process[18,32–35] in the framework of the transfer Hamiltonian formalism.[36,37] For this perspective we will focus on the latter as it allow us to derive a connection between the IET rate and the local density of optical states (LDOS) ρ_{opt} inside the tunnel gap.[35]

The efficiency of mode excitation by IET is determined by the ratio between the rate of inelastic tunneling Γ_{inel} and the rate of elastic tunneling Γ_{el}, which sum up to the total tunneling rate Γ as

$$\Gamma = \Gamma_{\mathrm{el}} + \Gamma_{\mathrm{inel}}. \tag{1}$$

Both of these contributions can be calculated in the framework of the transfer Hamiltonian formalism[36,37] that relies on the separation of the tunnel junction geometry into two subsystems, namely the two individual interfaces in the absence of the other electrode. Then, both elastic and inelastic tunneling can be introduced via perturbation theory. The elastic tunneling rate reads as

$$\Gamma_{\mathrm{el}} = \frac{2\pi}{\hbar} \int_0^{eV_{\mathrm{b}}} |\mathcal{T}(E)|^2 \, \rho_{\mathrm{L}}(E) \rho_{\mathrm{R}}(E) \, \mathrm{d}E, \tag{2}$$

whereas the spectral inelastic tunneling rate is given by[35]

$$\begin{aligned} \gamma_{\mathrm{inel}}(\hbar\omega) = {} & \frac{\pi e^2}{3\hbar\omega m^2 \varepsilon_0} \rho_{\mathrm{opt}} \\ & \times \int_{\hbar\omega}^{eV_{\mathrm{b}}} |\mathcal{P}(E, \hbar\omega)|^2 \, \rho_{\mathrm{R}}(E - \hbar\omega) \rho_{\mathrm{L}}(E) \, \mathrm{d}E, \end{aligned} \tag{3}$$

where $\mathcal{T}$ and $\mathcal{P}$ are the matrix elements for elastic and inelastic tunneling, respectively, $\rho_{\mathrm{L/R}}$ are the electronic densities of states of the left and right electrode, respectively, m is

the electron mass and ε_0 is the vacuum permittivity. The inelastic tunneling rate does not only depend on the electronic properties of the system but also on its optical properties, represented by the energy-dependent (partial) LDOS ρ_{opt}. The inelastic tunneling rate Γ_{inel} is obtained by integrating γ_{inel} over all energies as $\Gamma_{\mathrm{inel}} = \hbar \int_0^\infty \gamma_{\mathrm{inel}}(\hbar\omega)\mathrm{d}\omega$. For brevity, rates are stated in the limit $T \to 0$, where T is the temperature, reducing the Fermi distribution functions to step functions. The matrix elements are given by

$$\mathcal{T}(E) = \frac{\hbar^2}{2m}\left[\psi_{\mathrm{L}}\frac{\mathrm{d}\psi_{\mathrm{R}}^*}{\mathrm{d}z} - \psi_{\mathrm{R}}^*\frac{\mathrm{d}\psi_{\mathrm{L}}}{\mathrm{d}z}\right]_{z=z_0} \tag{4a}$$

$$\mathcal{P}(E,\hbar\omega) = -\,\mathrm{i}\hbar\int_0^d \psi_{\mathrm{R}}^*(E-\hbar\omega)\frac{\mathrm{d}}{\mathrm{d}z}\psi_{\mathrm{L}}(E)\,\mathrm{d}z. \tag{4b}$$

As no states are available inside the tunnel barrier, the wave functions of the two electrodes, $\psi_{\mathrm{L/R}}$, decay exponentially with distance from the respective electrode and are well described by

$$\psi_{\mathrm{L}}(z) = \psi_{\mathrm{L},0}\,\mathrm{e}^{-\kappa z}, \qquad z \geq 0 \tag{5a}$$

$$\psi_{\mathrm{R}}(z) = \psi_{\mathrm{R},0}\,\mathrm{e}^{-\kappa(d-z)}, \qquad z \leq d, \tag{5b}$$

with κ being the decay constant of the wave function and d being the thickness of the barrier.

The two transfer mechanisms are depicted in Fig. 2(a). Elastic tunneling connects electronic states of the same energy and occurs across the energy range given by the applied voltage as $|eV_{\mathrm{b}}|$, where occupied (left) and unoccupied (right) states overlap. Inelastic tunneling causes a transfer of electrons from an occupied state in the left electrode to an unoccupied state of lower energy in the right electrode. The energy range over which inelastic tunneling may occur is given by $|eV_{\mathrm{b}}| - \hbar\omega$. For the opposite bias polarity the roles of the electronic states are exchanged. For the sake of simplicity we may assume the electronic densities of states $\rho_{\mathrm{L/R}}$ and the decay constant κ to be independent of energy. Then, combining Eqs. (4)

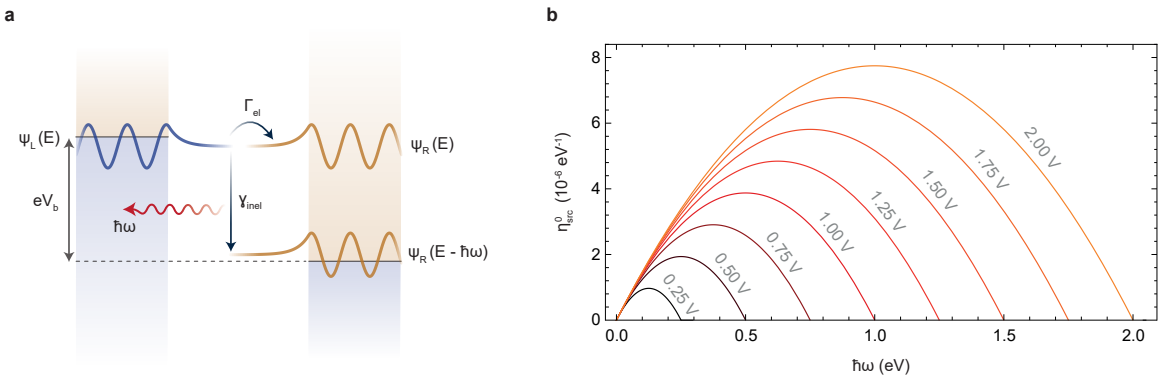

Figure 2: The source spectrum of IET. (a) Elastic and inelastic tunneling in the framework of the transfer Hamiltonian formalism. Elastic tunneling (Γ_{el}) is the result of the interaction between wave functions ψ_{L} and ψ_{R} of the same energy whereas inelastic tunneling (γ_{inel}) originates from the coupling between states with an energy difference given by the optical mode energy $\hbar\omega$. Inelastic tunneling is accompanied by the excitation of an optical mode. (b) IET (vacuum) source spectra $\eta^0_{\mathrm{src}}(\hbar\omega) = \gamma^0_{\mathrm{inel}}(\hbar\omega)/\Gamma_{\mathrm{el}}$, calculated for voltages ranging from 0.25 V to 2.0 V and for $d = 2\,\mathrm{nm}$.

and (5) allows us to express the matrix elements as

$$\mathcal{T} = \frac{\hbar^2\kappa}{m}\psi_{\mathrm{L},0}\,\psi_{\mathrm{R},0}\,\mathrm{e}^{-\kappa d} \tag{6a}$$

$$\mathcal{P} = -\mathrm{i}\hbar\kappa d\,\psi_{\mathrm{L},0}\,\psi_{\mathrm{R},0}\,\mathrm{e}^{-\kappa d}, \tag{6b}$$

which leads—in combination with Eqs. (2) and (3)—to the following expression for the source efficiency spectrum:

$$\begin{aligned}\eta_{\mathrm{src}}(\hbar\omega) &= \frac{\gamma_{\mathrm{inel}}(\hbar\omega)}{\Gamma_{\mathrm{el}}} \\ &= \frac{e^2d^2}{6\varepsilon_0\hbar^2\omega}\rho_0\left(1 - \frac{\hbar\omega}{eV_{\mathrm{b}}}\right)\frac{\rho_{\mathrm{opt}}}{\rho_0} \\ &= \eta^0_{\mathrm{src}}(\hbar\omega)\,\frac{\rho_{\mathrm{opt}}}{\rho_0},\end{aligned} \tag{7}$$

where $\rho_0 = \omega^2\pi^{-2}c^{-3}$ is the vacuum LDOS and c is the speed of light in vacuum.

Equation (7) allows us to analyze the importance of optical and electronic system properties separately. We introduced $\eta^0_{\mathrm{src}} = \gamma^0_{\mathrm{inel}}(\hbar\omega)/\Gamma_{\mathrm{el}}$ to describes the ratio of inelastic and elastic tunneling in vacuum, i.e. its efficiency. η^0_{src} is independent of the optical properties of the system and can be viewed as the source spectrum of IET. We plot η^0_{src} for voltages

ranging from 0.25 V to 2.0 V and for $d = 2\,\mathrm{nm}$ in Fig. 2(b). The applied bias determines the high energy cutoff of the spectrum as $\hbar\omega_{\max} = eV_{\mathrm{b}}$. Our calculations suggest that the efficiency of the inelastic process in vacuum is of the order of $10^{-6}\,\mathrm{eV}^{-1}$. Depending on the actual tunnel junction implementation, the assumption of constant electronic densities of states and energy-independent decay constant may not be justified and the respective dependencies need to be accounted for. For example, a rectangular barrier, assuming parabolic bands, leads to a source spectrum that is enhanced towards lower energies.[38]

Mode Excitation by IET in MIM Structures

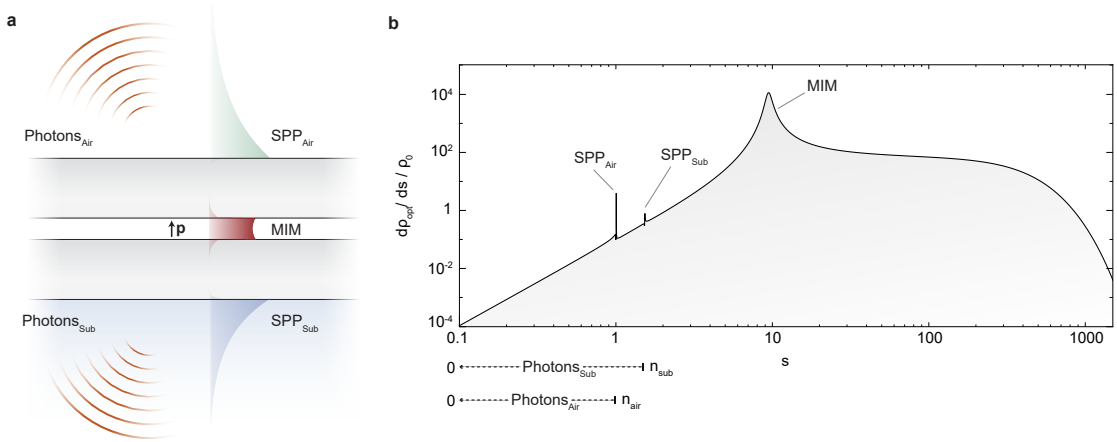

Figure 3: LDOS inside an MIM tunnel junction. (a) MIM tunnel junctions exhibit three guided modes that may be excited by a point dipole **p**: two SPP modes at the metal-air ($\mathrm{SPP_{Air}}$, green) and metal-substrate ($\mathrm{SPP_{Sub}}$, blue) interface, as well as one MIM mode (red) localized to the central insulating region. The modes are represented by their respective electric field (E_z) distribution (not to scale). Additionally, the geometry exhibits photon modes in the half-spaces above and below the MIM junction. (b) Spatial spectrum of the normalized LDOS in the center of the Al_2O_3 domain of a Glass ($\varepsilon_{\mathrm{Glass}} = (1.52)^2$)–Al (40 nm, $\varepsilon_{\mathrm{Al}}$ from ref 39)–Al_2O_3 (2 nm, $\varepsilon_{\mathrm{Al2O3}} = (1.75)^2$)–Au (40 nm, $\varepsilon_{\mathrm{Au}}$ from ref 40))–Air stack for $\hbar\omega = 1.0\,\mathrm{eV}$. The mode density associated with guided modes is given by the area enclosed by the corresponding peaks. The photon mode density is encoded in the spectral region $0 \leq s \leq n_{\mathrm{air/sub}}$, where $n_{\mathrm{air/sub}}$ is the refractive index of the air/substrate half-spaces (c.f. Supporting Information, Sect. S1).

So far we have established that IET can serve as a broadband excitation source for electromagnetic modes. Equation (7) states that the LDOS determines the efficiency with

which a certain mode is excited. In order to determine the LDOS inside the tunneling gap, we make use of the fact that the normalized LDOS $\rho_{\mathrm{opt}}/\rho_0$ is equivalent to the normalized power P/P_0 dissipated by a point dipole of moment $\mathbf{p}$ (c.f. Fig. 3(a)) as[41]

$$\frac{\rho_{\mathrm{opt}}}{\rho_0} = \frac{P}{P_0}, \tag{8}$$

where $P_0 = \omega^4|\mathbf{p}|/(12\pi\varepsilon_0 c^3)$ is the power radiated by the dipole in vacuum. The dipole orientation is defined by the direction of electron tunneling. The power dissipated by the dipole is

$$P = \frac{\omega}{2}\, p\, \mathrm{Im}\left\{E_z(\mathbf{r}_0)\right\}, \tag{9}$$

where $\mathbf{r}_0$ is the position of the dipole and E_z is the dipole's electric field projected on the normal axis z, that is, perpendicular to the interfaces. For the case of an MIM junction, or generally for any layered medium, the latter can be expressed as a superposition of plane and evanescent waves of varying in-plane wave vector $k_\parallel$ and their interaction with the environment can be treated separately for each evanescent / plane wave. The angular spectrum representation of the field reads as[41]

$$E_z(z, z_0) = \frac{\mathrm{i}\omega^3 p}{4\pi\varepsilon_0\varepsilon_i c^3} \int_0^\infty \frac{s^3}{s_{zi}} \mathrm{e}^{\mathrm{i}k_0 s_{zi}|z-z_0|}\, \mathrm{d}s, \tag{10}$$

where ε_i is the dielectric permittivity of the medium in which the dipole is located, $s = k_\parallel/k_0$ is the normalized in-plane wave vector component with k_0 being the wave vector in vacuum and $s_{zi} = k_{zi}/k_0 = \sqrt{\varepsilon_i - s^2}$ is the normalized out-of-plane wave vector component. For a dipole embedded in a layered medium consisting of N layers, situated in layer i at the position z_0, the field generated by each plane or evanescent wave can be expressed as[42]

$$\mathcal{F}_j\left(z, z_0\right) = \delta_{ij}\mathrm{e}^{\mathrm{i}k_0 s_{zi}|z-z_0|} + c_j^{\uparrow}\mathrm{e}^{\mathrm{i}k_0 s_{zj} z} + c_j^{\downarrow}\mathrm{e}^{-\mathrm{i}k_0 s_{zj} z}, \tag{11}$$

where $c_j^{\uparrow}$ and $c_j^{\downarrow}$ are complex field amplitudes of upward and downward propagating waves, respectively. Consequently, the field generated by the dipole at its own position $z = z_0 = 0$, follows from Eqs. (10) and (11) and reads as

$$E_z = \frac{\mathrm{i}\omega^3 p_z}{4\pi\varepsilon_0\varepsilon c^3} \int_0^\infty \frac{s^3}{s_{zi}} \left(1 + c_i^{\uparrow} + c_i^{\downarrow}\right) \mathrm{d}s, \tag{12}$$

which in combination with Eq. (9) yields the following expression for the normalized dissipated power or, equivalently, the normalized LDOS:

$$\frac{P}{P_0} = \frac{\rho_{\mathrm{opt}}}{\rho_0} = \frac{3}{2} \int_0^\infty \mathrm{Re}\left\{ \frac{s^3}{s_{zi}\varepsilon_i} \left(1 + c_i^{\uparrow} + c_i^{\downarrow}\right) \right\} \mathrm{d}s. \tag{13}$$

The complex field amplitudes $c_j^{\uparrow}$ and $c_j^{\downarrow}$ are determined by the boundary conditions for p-polarized fields following the ansatz given by Eq. (11).[41] The advantage of determining the LDOS following Eq. (13) compared to numerical means is that it allows us to analyze the spatial spectrum of the LDOS and determine the LDOS distribution across the different available modes.

For this perspective we focus on one particular MIM junction configuration, namely Aluminum (Al)–Aluminum Oxide (Al_2O_3)–Gold (Au), since it has been one of the most commonly used configurations from the original work by Lambe & McCarthy[3] until today. Exemplarily, for one particular mode energy ($\hbar\omega = 1\,\mathrm{eV}$), we show the spatial spectrum of the LDOS in the center of the tunnel barrier in Fig. 3(b). Superimposed to a broad 'background' we find three peaks that correspond to the three modes of the geometry. These modes are an SPP mode bound to the Au–Air interface, an SPP mode bound to the Al–Glass interface and the MIM mode, as illustrated in Fig. 3(a). Integrating the spatial spectrum over all frequencies yields the total LDOS ρ_{tot} associated with the normal axis. Here we find that $\rho_{\mathrm{tot}} \sim 4 \times 10^4\,\rho_0$, which means that the probability for IET, compared to vacuum, is increased by more than four orders of magnitude. But how is this mode density distributed

across the different modes of the geometry? The area enclosed by a peak corresponds to the LDOS associated with the respective mode. For a mode energy of 1 eV we find that close to 50 % of ρ_{tot} is associated with the MIM mode. The LDOS fraction associated with the two SPP modes on the other hand is of the order of 10^{-7}. In addition to the coupling to bound modes, the spatial spectrum further reveals the coupling to the radiation continuum, i.e. the emission of photons. The associated LDOS fraction is of the order of 2.5×10^{-8} (c.f. Supporting Information, Sect. S1). The remaining 50 % of ρ_{tot} is due to the broad background which is caused by quenching, i.e. the direct dissipation of electromagnetic energy. It is thus apparent that IET predominantly drives the MIM mode.

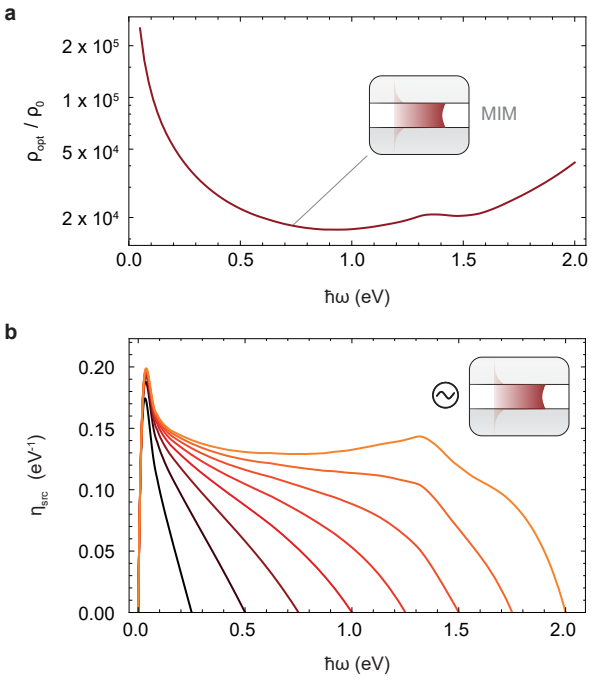

Figure 4: The source efficiency of IET in an MIM tunnel junction. (a) Normalized LDOS associated with the MIM mode as a function of energy $\hbar\omega$. The values have been extracted as the areas of the MIM peaks in the spatial spectra (c.f. Fig. 3(b) and Supporting Information, Fig. S1). (b) Source efficiency spectrum $\eta_{\mathrm{src}} = \eta^0_{\mathrm{src}} \times \rho_{\mathrm{opt}}/\rho_0$ (c.f. Eq. (7)), for voltages ranging from 0.25 V to 2.0 V, with $\rho_{\mathrm{opt}}/\rho_0$ from (a) and η^0_{src} from Fig. 2(b).

In Fig. 4(a) we show the LDOS of the MIM mode as a function of mode energy. The LDOS varies between 10^4 and $10^5\,\rho_0$. The origin of this extreme enhancement lies in its high effective mode index as well as the mode confinement by the two metallic electrodes.[43]

Equation (7) allows us to calculate the source efficiency spectrum of IET, which is shown in Fig. 4(b). Our analysis shows that the probability for a tunneling electron to excite the MIM mode is of the order of $0.1\,\mathrm{eV}^{-1}$. Spectral integration yields an overall efficiency of the order of 10 %, in excellent agreement with the early theoretical predictions of L. C. Davis.[25]

Mode Propagation and Outcoupling

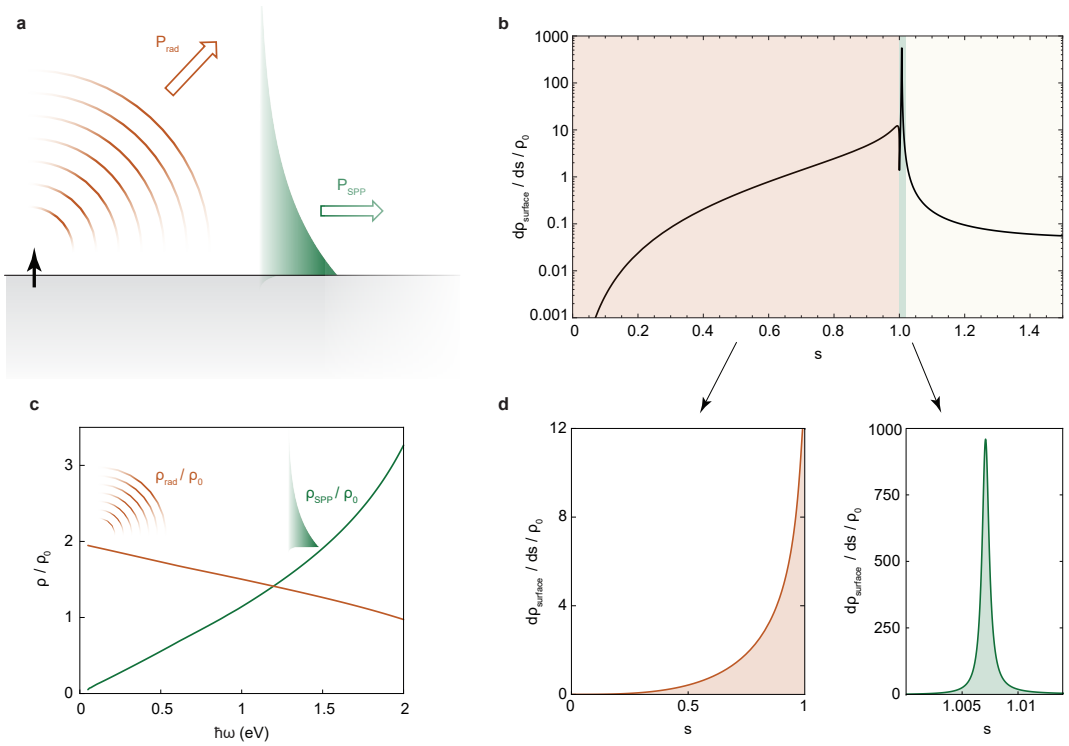

Figure 5: Coupling to radiation (photons) and SPPs. (a) The MIM mode energy can be coupled to radiation, to SPPs, or absorbed in the metal. The dissipated power deelecpends on the LDOS near the metal surface. (b) Spatial spectrum of the LDOS at the surface of gold for $\hbar\omega = 1\,\mathrm{eV}$. The regions of radiation (red) and SPPs (green) are shown separately on a linear scale in (d). (c) Normalized power coupled to radiation ($\rho_{\mathrm{rad}}/\rho_0$) and SPPs ($\rho_{\mathrm{SPP}}/\rho_0$) as a function of energy, obtained from the areas enclosed by the spatial spectra shown in (d). The energy coupled to radiation and to SPPs is roughly equal.

Our analysis shows that IET efficiently drives the MIM mode as the LDOS associated with the MIM mode is able to compensate for the intrinsically low IET source efficiency. So why have electron to photon conversion efficiencies remained comparatively low despite

decades of research? The difficulty lies in the conversion of the MIM mode to the desired output mode, i.e. photons or SPPs (c.f. Fig. 1). The confinement of the MIM mode to the nanoscale gap is accompanied by large propagation losses. The propagation length of the mode is determined by the imaginary part of the propagation constant as $L = 1/(2\,\mathrm{Im}(k_\parallel))$, which in turn is related to the full-width at half-maximum (FWHM) of the MIM peaks in the spatial spectrum of the LDOS (c.f. Fig. 3(b)) as $\mathrm{Im}(k_\parallel) = k_0 \times \mathrm{FWHM}$. We find that the propagation length of the MIM mode decreases from $L(\hbar\omega = 0.05\,\mathrm{eV}) \sim 1\,\mu\mathrm{m}$ to $L(\hbar\omega = 2\,\mathrm{eV}) \sim 20\,\mathrm{nm}$ (c.f. Supporting Information, Fig. S2).

In the initial series of experiments in the years following the discovery of light emission from IET, the lateral dimensions of MIM devices were several orders of magnitude larger than the propagation length of the MIM mode. As the propagation constant of the MIM mode is approximately one order of magnitude higher than the propagation constant of photons or SPPs (c.f. Fig. 3(b) and Supporting Information, Sect. S2) a direct coupling between the MIM mode and the other modes of the geometry is not possible. Hence, in an experimental realization of the (laterally) macroscopic structure shown in Fig. 3(a), one would merely observe the emission of photons and SPPs that are directly excited by IET. Due to the low LDOS of these modes inside the MIM junction, the efficiency of SPP/photon generation would be of the order of $\sim 10^{-8}/10^{-9}$ (c.f. Fig. 3(b) and Supporting Information, Fig. S1). Several strategies were followed in order to convert the MIM mode into photons in these early experiments. Amongst these were the roughening of the junction electrodes[44–50] and the use of metal nanoparticles,[44,46,51–53] both of which facilitate mode conversion through random scattering. However, as each scattering event carries the possibility for absorption, outcoupling efficiencies were found to saturate with increasing roughness.[4,44,54] In an alternative approach tunnel junctions were fabricated on top of gratings[55–59] in order to 'impedance-match' the modes of the tunnel junctions to free-space (c.f. Fig. 3(a)). Both approaches are limited as they rely on mode conversion through the metallic electrodes which leads to considerable losses.

A more promising approach is to convert the MIM mode into the output mode by terminating the MIM gap. In this approach, there are two main factors that limit the efficiency: propagation losses and coupling losses. We first discuss the former and then the latter.

The limited propagation length of the MIM mode, discussed previously, sets stringent requirements on the lateral size of the MIM device. For devices significantly larger than the propagation length only a small fraction of the total MIM mode energy will reach the end facets of the device.

Next, the MIM mode energy that does reach the end facets needs to be coupled to the desired output mode, either free radiation (photons) or SPPs. The large propagation constant of the MIM mode compared to the propagation constant of photons or SPPs (c.f. Fig. 5(b)) further complicates the conversion. This 'impedance mismatch' causes MIM modes to be reflected from the end facets of the MIM device. The outcoupling efficiency depends on the properties of the 'coupling network' between the source (MIM mode) and the output (photons or SPPs). A coupling network can be designed to resonantly enhance the efficiency in a desired band of temporal frequencies $\Delta\omega$ and/or spatial frequencies $\Delta k_{\|}$, as has been shown in recent work by means of optical antennas.[14,15]

To analyze whether there is any fundamental advantage of the SPP output compared to the photon output, we consider the case where no resonant enhancement is used in either the temporal or the spatial frequency domain, i.e. the direct coupling between the MIM device and a planar metal surface. This scenario corresponds to the configuration shown in Fig. 1 with the gap closed between source and output, that is, with the metal surface directly connected to the MIM end facet. The impedance mismatch between the MIM mode and both output modes leads to high reflection losses at the interface. The outcoupling efficiency now depends on the LDOS of the planar metal surface (ρ_{surface}) and is the highest right at the surface of the metal. It can be represented as[41]

$$\rho_{\mathrm{surface}} \;=\; \frac{3\omega^2}{2\pi^2 c^3}\,\mathrm{Re}\int_0^\infty \frac{s^3}{\sqrt{1-s^2}}\,\left[1 \,+\, r^p(s)\right]\,\mathrm{d}s\;, \tag{14}$$

where $s = k_\parallel/k_0$ is the normalized in-plane wavevector component ($k_0 = \omega/c$) and r^p is the Fresnel reflection coefficient for p polarization. For $r^p = 0$, Eq. (14) evaluates to $\rho_0 = \omega^2/(\pi^2 c^3)$, in agreement with the LDOS of free space.

In Fig. 5(b) we show the spatial spectrum of ρ_{surface} for a gold surface. It features a distinctive peak at $k_\parallel = k_{\text{SPP}}$, which is associated with the SPP mode. To determine the fraction of energy that can be coupled to the SPP mode we integrate the spatial spectrum over the width of the SPP peak. Similarly, the fraction of energy coupled to photons is calculated by integrating the LDOS spectrum over the range of real k-vectors, that is, $k_\parallel = [0\,..\,k_0]$. The result is shown in Fig. 5(c) as a function of energy $\hbar\omega$. At the termination of the MIM gap, the transmission into each mode will be proportional to its LDOS. We find that the amount of energy coupled to photons and SPPs is roughly the same. Hence, in the absence of coupling elements that are specifically designed to favor either output mode, coupling to SPPs rather than photons bears no significant advantage. Consequently, in this case, the efficiency of SPP excitation by IET cannot be significantly higher than the efficiency of photon excitation.

Design rules for efficient photon or SPP sources

The external device efficiency η_{ext}, defined as the number of emitted photons (or SPPs) per tunneling electron and per unit spectral width, is given by

$$\eta_{\text{ext}} = \eta^0_{\text{src}} \times \frac{\rho_{\text{opt}}}{\rho_0} \times \eta_{\text{em}} = \eta_{\text{src}} \times \eta_{\text{em}}\,, \tag{15}$$

with η_{src} being the efficiency of electron to MIM mode conversion and η_{em} being the efficiency of converting the MIM mode to photons or SPPs. In an MIM tunnel junction, η_{src} is of the order of $0.1\,\text{eV}^{-1}$ (c.f. Fig. 4(b)), enabled by the extremely high LDOS of the MIM mode, suggesting that an efficient photon/SPP source based on IET is indeed feasible. However, irrespective of the final output, the decisive factor that ultimately determines the external

device efficiency is η_{em}. Two design rules follow immediately from the previous discussion.

1. An efficient conversion is only possible if the lateral junction dimensions are not significantly larger than the MIM mode propagation length.

2. The impedance mismatch between the MIM mode and the output needs to be overcome by a properly designed outcoupling element, e.g. optical antenna.

While these rules appear straightforward at first sight, their experimental implementation has challenged researchers for several decades.

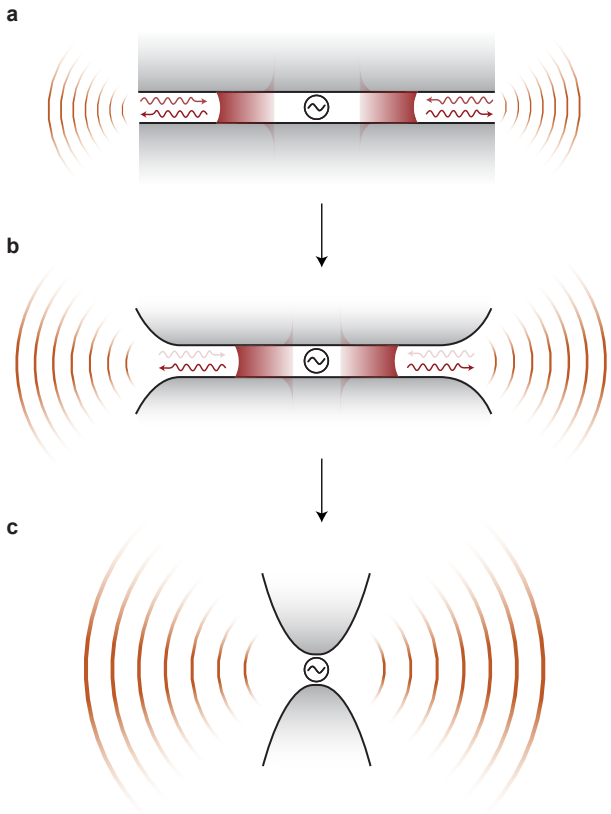

Figure 6: Design rules for efficient photon sources. (a) IET (source symbol) drives the MIM mode, which propagates towards the gap termination. The majority of MIM SPPs is reflected at the facet due to the large impedance mismatch with free space radiation. (b) By adiabatically opening the ends of the MIM gap we achieve gradual wave vector changes which leads to strongly reduced reflection. (c) A shortened MIM gap further reduces propagation losses.

The illustrations of Fig. 6 visualize the implementation of these rules, leading to a design that is reminiscent of a bow-tie antenna. In antenna-coupled tunnel junctions, the role of

the antenna is to optimize η_{em}. Here, optical antennas serve as the element connecting the two half-spaces in the illustration of Fig. 1(a), facilitating the conversion of the MIM mode inside the tunnel gap into photons or SPPs.[1,14,15,18–20,22–24,35] Optical antennas also enhance the source efficiency through the Purcell effect. The geometry studied by Qian et al. employs silver nanocrystals aligned edge to edge in order to form the tunnel junction.[1] As the nanocrystals are tens of nanometers in size and consist of single-crystalline silver, propagation losses are reduced to a minimum. Furthermore, by varying size and aspect ratio of the nanocrystals Qian et al. spectrally overlap resonances in both the LDOS ($\rho_{\mathrm{opt}}/\rho_0$) and the radiation efficiency (η_{em}). This combination allowed them to reach electron to photon conversion efficiencies as high as 2 %.

The on-chip SPP source presented by Du et al.[2] is reminiscent of the very first realizations of IET-based light sources, consisting of an oxidized aluminum thin-film crossed by a gold counter electrode. Du et al. argue that their devices are efficient SPP sources as their devices are not limited by low photon outcoupling efficiencies. However, IET predominantly excites the MIM mode, which needs to be converted to the desired output mode. For MIM junctions that are directly coupled to planar metal surfaces, η_{em} is expected to be low and was shown to be roughly the same for SPPs and photons. Thus, there is no immediate advantage in converting MIM modes to SPPs as opposed to photons. We further note that the lateral dimensions of the MIM devices studied by Du et al. are significantly larger than the propagation length of MIM modes. Hence, most of the energy contained in MIM modes is absorbed in the tunnel gap. These observations, in conjunction with numerical simulations presented in Sect. S3 of the Supporting Information, lead us to conclude that the excitation of SPPs by IET is very unlikely to explain the measurements by Du et al.

Conclusion

In conclusion, we have presented a theoretical analysis of SPP and photon generation by IET in MIM device geometries. In agreement with earlier theoretical works we showed that IET efficiently excites MIM modes due to LDOS enhancement. We identified the conversion of the MIM mode to either photons or SPPs as the primary challenge in the realization of efficient SPP or photon sources and discussed strategies to facilitate this conversion. Based on the progress that has been made in recent years and the largely untapped potential of novel materials and device geometries we are convinced that there is light at the end of the tunnel for IET-based photon and SPP sources.

Acknowledgement

The authors thank M. Frimmer for valuable discussions. This work was supported by the Swiss National Science Foundation (grant no. 200021_165841) and the ETH Zürich (ETH-32 15-1).

Supporting Information Available

Details of the LDOS inside an MIM tunnel junction including the energy-dependence of the total LDOS, the guide mode LDOS and the photon LDOS; equations describing the photon emission into the lower and upper half-space; dispersion relation (propagation constant and propagation length) of modes supported by MIM devices; numerical (finite element) simulations of the spatial LDOS variation in the source geometry of ref 2.

References

(1) Qian, H.; Hsu, S.-W.; Gurunatha, K.; Riley, C. T.; Zhao, J.; Lu, D.; Tao, A. R.; Liu, Z. Efficient light generation from enhanced inelastic electron tunnelling. *Nature Photon.* **2018**, *12*, 485–488.

(2) Du, W.; Wang, T.; Chu, H.-S.; Nijhuis, C. A. Highly efficient on-chip direct electronic–plasmonic transducers. *Nature Photon.* **2017**, *11*, 623.

(3) Lambe, J.; McCarthy, S. L. Light emission from inelastic electron tunneling. *Phys. Rev. Lett.* **1976**, *37*, 923–925.

(4) Mills, D.; Weber, M.; Laks, B. In *Tunneling spectroscopy: capabilities, applications, and new techniques*; Hansma, P. K., Ed.; Plenum Press: New York, 1982; Chapter 5, pp 121–152.

(5) Coombs, J. H.; Gimzewski, J. K.; Reihl, B.; Sass, J. K.; Schlittler, R. R. Photon emission experiments with the scanning tunnelling microscope. *J. Microsc.* **1988**, *152*, 325–336.

(6) Gimzewski, J. K.; Reihl, B.; Coombs, J. H.; Schlittler, R. R. Photon emission with the scanning tunneling microscope. *Z. Phys. B* **1988**, *72*, 497–501.

(7) Berndt, R. In *Scanning Probe Microscopy: Analytical Methods*; Wiesendanger, R., Ed.; Springer: Berlin Heidelberg, 1998; pp 97–134.

(8) Sakurai, M.; Thirstrup, C.; Aono, M. New aspects of light emission from STM. *Appl. Phys. A* **2005**, *80*, 1153–1160.

(9) Rossel, F.; Pivetta, M.; Schneider, W.-D. Luminescence experiments on supported molecules with the scanning tunneling microscope. *Surf. Sci. Rep.* **2010**, *65*, 129–144.

(10) Bharadwaj, P.; Deutsch, B.; Novotny, L. Optical antennas. *Adv. Opt. Phot.* **2009**, *1*, 438–483.

(11) Biagioni, P.; Huang, J.-S.; Hecht, B. Nanoantennas for visible and infrared radiation. *Rep. Prog. Phys.* **2012**, *75*, 024402.

(12) Bharadwaj, P.; Bouhelier, A.; Novotny, L. Electrical excitation of surface plasmons. *Phys. Rev. Lett.* **2011**, *106*, 226802.

(13) Le Moal, E.; Marguet, S.; Rogez, B.; Mukherjee, S.; Dos Santos, P.; Boer-Duchemin, E.; Comtet, G.; Dujardin, G. An electrically excited nanoscale light source with active angular control of the emitted light. *Nano Lett.* **2013**, *13*, 4198–4205.

(14) Kern, J.; Kullock, R.; Prangsma, J.; Emmerling, M.; Kamp, M.; Hecht, B. Electrically driven optical antennas. *Nature Photon.* **2015**, *9*, 582–586.

(15) Parzefall, M.; Bharadwaj, P.; Jain, A.; Taniguchi, T.; Watanabe, K.; Novotny, L. Antenna-coupled photon emission from hexagonal boron nitride tunnel junctions. *Nature Nanotech.* **2015**, *10*, 1058–1063.

(16) Du, W.; Wang, T.; Chu, H.-S.; Wu, L.; Liu, R.; Sun, S.; Phua, W. K.; Wang, L.; Tomczak, N.; Nijhuis, C. A. On-chip molecular electronic plasmon sources based on self-assembled monolayer tunnel junctions. *Nature Photon.* **2016**, *10*, 274–280.

(17) Dathe, A.; Ziegler, M.; Hübner, U.; Fritzsche, W.; Stranik, O. Electrically Excited Plasmonic Nanoruler for Biomolecule Detection. *Nano Lett.* **2016**, *16*, 5728–5736.

(18) Uskov, A. V.; Khurgin, J. B.; Protsenko, I. E.; Smetanin, I. V.; Bouhelier, A. Excitation of plasmonic nanoantennas with nonresonant and resonant electron tunnelling. *Nanoscale* **2016**, *8*, 14573.

(19) Vardi, Y.; Cohen-Hoshen, E.; Shalem, G.; Bar-Joseph, I. Fano Resonance in an Electrically Driven Plasmonic Device. *Nano Lett.* **2016**, *16*, 748–752.

(20) Bigourdan, F.; Hugonin, J.-P.; Marquier, F.; Sauvan, C.; Greffet, J.-J. Nanoantenna for electrical generation of surface plasmon polaritons. *Phys. Rev. Lett.* **2016**, *116*, 106803.

(21) Qian, H.; Hsu, S.-W.; Gurunatha, K.; Zhao, J.; Riley, C. T.; Lu, D.; Tao, A.; Liu, Z. Investigation of the light generation from crystalline Ag-cubes based metal-insulator-metal tunnel junctions. CLEO: QELS_Fundamental Science. 2017; pp FM3H–7.

(22) Gurunarayanan, S. P.; Verellen, N.; Zharinov, V. S.; James Shirley, F.; Moshchalkov, V. V.; Heyns, M.; Van de Vondel, J.; Radu, I. P.; Van Dorpe, P. Electrically driven unidirectional optical nanoantennas. *Nano Lett.* **2017**, *17*, 7433–7439.

(23) Namgung, S.; Mohr, D. A.; Yoo, D.; Bharadwaj, P.; Koester, S. J.; Oh, S.-H. Ultrasmall Plasmonic Single Nanoparticle Light Source Driven by a Graphene Tunnel Junction. *ACS Nano* **2018**, *12*, 2780–2788.

(24) Parzefall, M.; Szabó, Á.; Taniguchi, T.; Watanabe, K.; Luisier, M.; Novotny, L. Light from Van der Waals quantum tunneling devices. *arXiv:1804.06163 [cond-mat.mes-hall]* **2018**,

(25) Davis, L. C. Theory of surface-plasmon excitation in metal-insulator-metal tunnel junctions. *Phys. Rev. B* **1977**, *16*, 2482–2490.

(26) Johansson, P.; Monreal, R.; Apell, P. Theory for light emission from a scanning tunneling microscope. *Phys. Rev. B* **1990**, *42*, 9210–9213.

(27) Johansson, P. Light emission from a scanning tunneling microscope: Fully retarded calculation. *Phys. Rev. B* **1998**, *58*, 10823–10834.

(28) Hone, D.; Mühlschlegel, B.; Scalapino, D. J. Theory of light emission from small particle tunnel junctions. *Appl. Phys. Lett.* **1978**, *33*, 203–204.

(29) Rendell, R. W.; Scalapino, D. J.; Mühlschlegel, B. Role of Local Plasmon Modes in Light Emission from Small-particle Tunnel Junctions. *Phys. Rev. Lett.* **1978**, *41*, 1746–1750.

(30) Rendell, R. W.; Scalapino, D. J. Surface plasmons confined by microstructures on tunnel junctions. *Phys. Rev. B* **1981**, *24*, 3276–3294.

(31) Uehara, Y.; Kimura, Y.; Ushioda, S.; Takeuchi, K. Theory of visible light emission from scanning tunneling microscope. *Jpn. J. Appl. Phys.* **1992**, *31*, 2465–2469.

(32) Persson, B. N. J.; Baratoff, A. Theory of photon emission in electron tunneling to metallic particles. *Phys. Rev. Lett.* **1992**, *68*, 3224–3227.

(33) Schimizu, T.; Kobayashi, K.; Tsukada, M. Theoretical approach to the microscopic mechanism of light emission from a scanning tunneling microscope. *Appl. Surf. Sci.* **1992**, *60-61*, 454–459.

(34) Downes, A.; Taylor, M. E.; Welland, M. E. Two-sphere model of photon emission from the scanning tunneling microscope. *Phys. Rev. B* **1998**, *57*, 6706–6714.

(35) Parzefall, M.; Bharadwaj, P.; Novotny, L. In *Quantum Plasmonics*; Bozhevolnyi, S. I., Martin-Moreno, L., Garcia-Vidal, F., Eds.; Springer: Cham, 2017; pp 211–236.

(36) Bardeen, J. Tunnelling from a many-particle point of view. *Phys. Rev. Lett.* **1961**, *6*, 57–59.

(37) Duke, C. B. *Tunneling in solids*; Academic Press: New York, 1969; Vol. 10.

(38) Parzefall, M. Optical antennas driven by quantum tunneling. Ph.D. thesis, ETH Zurich, 2017.

(39) Rakić, A. D. Algorithm for the determination of intrinsic optical constants of metal films: application to aluminum. *Appl. Opt.* **1995**, *34*, 4755–4767.

(40) Olmon, R. L.; Slovick, B.; Johnson, T. W.; Shelton, D.; Oh, S.-H.; Boreman, G. D.; Raschke, M. B. Optical dielectric function of gold. *Phys. Rev. B* **2012**, *86*, 235147.

(41) Novotny, L.; Hecht, B. *Principles of Nano-Optics*, 2nd ed.; Cambridge University Press: Cambridge, 2012.

(42) Chew, W. C. *Waves and Fields in Inhomogeneous Media*, 1st ed.; Van Nostrand Reinhold: New York, 1990.

(43) Faggiani, R.; Yang, J.; Lalanne, P. Quenching, plasmonic, and radiative decays in nanogap emitting devices. *ACS Photonics* **2015**, *2*, 1739–1744.

(44) McCarthy, S. L.; Lambe, J. Enhancement of light emission from metal-insulator-metal tunnel junctions. *Appl. Phys. Lett.* **1977**, *30*, 427–429.

(45) Jain, R. K.; Wagner, S.; Olson, D. H. Stable room-temperature light emission from metal-insulator-metal junctions. *Appl. Phys. Lett.* **1978**, *32*, 62–64.

(46) Adams, A.; Hansma, P. K. Light emission from small metal particles and thin metal films excited by tunneling electrons. *Phys. Rev. B* **1981**, *23*, 3597–3601.

(47) Parvin, K.; Parker, W. Optical spectra and angular dependence of the visible light emitted by metal-insulator-metal tunnel junctions. *Solid State Comm.* **1981**, *37*, 629–633.

(48) Walmsley, D. G.; Quinn, H. F.; Dawson, P. Image of the electron energy-loss function in light emitted from tunnel junctions. *Phys. Rev. Lett.* **1982**, *49*, 892–895.

(49) Dawson, P.; Walmsley, D. G.; Quinn, H. A.; Ferguson, A. J. L. Observation and explanation of light-emission spectra from statistically rough Cu, Ag and Au tunnel junctions. *Phys. Rev. B* **1984**, *30*, 3164–3178.

(50) Hänisch, M.; Otto, A. Light emission from rough tunnel junctions in UHV. *J. Phys.: Condens. Matter* **1994**, *6*, 9659–9676.

(51) Hansma, P. K.; Broida, H. P. Light emission from gold particles excited by electron tunneling. *Appl. Phys. Lett.* **1978**, *32*, 545–547.

(52) Adams, A.; Wyss, J. C.; Hansma, P. K. Possible observation of local plasmon modes excited by electrons tunneling through junctions. *Phys. Rev. Lett.* **1979**, *42*, 912–915.

(53) Bloemer, M. J.; Mantovani, J. G.; Goudonnet, J. P.; James, D. R.; Warmack, R. J.; Ferrell, T. L. Observation of driven surface-plasmon modes in metal particulates above tunnel junctions. *Phys. Rev. B* **1987**, *35*, 5947–5954.

(54) Laks, B.; Mills, D. L. Light emission from tunnel junctions: the role of the fast surface polariton. *Phys. Rev. B* **1980**, *22*, 5723–5729.

(55) Kirtley, J. R.; Theis, T. N.; Tsang, J. C. Diffraction-grating-enhanced light emission from tunnel junctions. *Appl. Phys. Lett.* **1980**, *37*, 435–437.

(56) Kirtley, J.; Theis, T. N.; Tsang, J. C. Light emission from tunnel junctions on gratings. *Phys. Rev. B* **1981**, *24*, 5650–5663.

(57) Kroo, N.; Szentirmay, Z.; Felszerfalvi, J. On the origin of light emission by tunnel junctions. *Phys. Lett. A* **1981**, *81*, 399–401.

(58) Sparks, P. D.; Sjodin, T.; Reed, B. W.; Stege, J. Light emission from the slow mode of tunnel junctions on short period diffraction gratings. *Phys. Rev. Lett.* **1992**, *68*, 2668–2671.

(59) Ushioda, S.; Uehara, Y.; Takada, M.; Otsubo, K.; Murota, J. Grating-coupled light emission from the slow mode of metal-insulator-metal tunnel junctions. *Jpn. J. Appl. Phys.* **1992**, *31*, L870–L873.

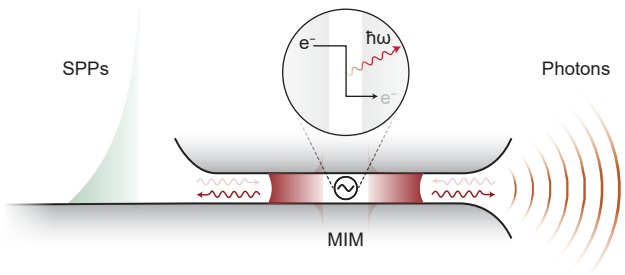

Figure 7: Table of Contents Graphic