Plasmon generation through electron tunneling in twisted double-layer graphene and metal-insulator-graphene systems

Sandra de Vega[1] and F. Javier García de Abajo[1,2,*]

[1]*Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[2]*Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys, 23, 08010 Barcelona, Spain*

The generation of highly confined plasmons through far-field optical illumination appears to be impractical for technological applications due to their large energy-momentum mismatch with external light. Electrical generation of plasmons offers a possible solution to this problem, although its performance depends on a careful choice of material and geometrical parameters. Here we theoretically investigate graphene-based structures and show in particular the very different performances between (i) two layers of graphene separated by a dielectric and (ii) metal/insulator/graphene sandwiches as generators of propagating plasmons assisted by inelastic electron tunneling. For double-layer graphene, we study the dependence on the relative tilt angle between the two sheets and show that the plasmon generation efficiency for 4° twist angle drops to ∼20% from its maximum for perfect stacking. For metal/insulator/graphene sandwiches, the inelastic tunneling efficiency drops by several orders of magnitude relative to double-layer graphene, regardless of doping level, metal/graphene separation, choice of metal, and direction of tunneling (metal to or from graphene), a result that we attribute to the small fraction of the surface-projected metal Brillouin zone covered by the graphene Dirac cone. Our results reveal a reasonable tolerance to graphene lattice misalignment and a poor performance of structures involving metals, thus supporting the use of double-layer graphene as an optimum choice for electrical plasmon generation in tunneling devices.

I. INTRODUCTION

Surface plasmons, the collective excitations of conduction electrons at the surface of conducting materials, have been the focus of intense research over the past decades because of their ability to interact strongly with light, producing a huge concentration of electromagnetic energy down to deep-subwavelength regions, accompanied by strong enhancement of the optical field intensity [1,2]. Plasmons are relatively tolerant to defects in the fabrication of the structures supporting them, which commonly rely on colloid chemistry [3] and lithography [4] techniques, with application to optical sensing [5], photocatalysis [6], nonlinear optics [7], and optical signal processing [8], among other feats. However, the strong spatial confinement produced by optical plasmons has a negative side effect: the coupling cross section from far-field radiation to plasmons is small, typically comparable to or even smaller than the lateral size of the supporting plasmonic nanostructures. Plasmon excitation with light is thus inefficient, which represents a serious obstacle in the development of practical applications. Various methods have been devised to solve this problem, among them the use of funneling structures that focus light down to the spatial extension of the plasmon through the use of gratings [9,10] and tips [11–13]. Quantum emitters have also been employed as a source of plasmons [14], and although they can operate as a source of single plasmons, their efficiency is also low.

An alternative to the optical generation of plasmons is provided by electron beams, which have been used in pioneering studies of surface plasmons, including their discovery [15–17]. However, the production and control of electron beams are difficult to combine with integrated devices. Electrons can also inelastically undergo tunneling between neighboring conductors, losing energy that results in the emission of light [18,19]. Recently, plasmon emission produced by inelastic tunneling has been explored in different systems [20,21], and this mechanism has also been theoretically explored in sandwiches formed between graphene layers [22–25], where it is found to be particularly efficient [24].

In this paper, we extend a previous work [24] and present a comprehensive study of the plasmon-emission efficiency associated with inelastic tunneling between two graphene layers. Also, we compare the results with the emission in metal/insulator/graphene (MIG) configurations. For double-layer graphene (DLG), we assess the dependence of the emission efficiency on the relative graphene twisting angle and find a substantial reduction in the efficiency when the K points in the two layers are misaligned by more than a few degrees. For MIG configurations, we find plasmon generation rates way below those of DLG.

II. SPECTRALLY RESOLVED INELASTIC TUNNELING CURRENT

The large plasmon confinement and small thickness compared with the light wavelength in the structures under consideration allow us to work within the quasistatic limit. We thus calculate the probability for an electron to inelastically tunnel

*javier.garciadeabajo@nanophotonics.es

between initial and final states ψ_i and ψ_f while transferring an amount of energy $\hbar\omega$ to the materials (e.g., a plasmon) by using the frequency-resolved screened interaction $W(\mathbf{r}, \mathbf{r}', \omega)$, which describes the potential created at $\mathbf{r}$ by a point charge placed at $\mathbf{r}'$ and oscillating with frequency ω. More precisely, the probability can be written as [17,24]

$$
\Gamma(\omega) = 4\frac{2e^2}{\hbar} \sum_{i,f} \int d^3\mathbf{r} \int d^3\mathbf{r}'\, \psi_i^\dagger(\mathbf{r}) \cdot \psi_f(\mathbf{r})\psi_f^\dagger(\mathbf{r}') \cdot \psi_i(\mathbf{r}')
$$
$$
\times \operatorname{Im}\{-W(\mathbf{r}, \mathbf{r}', \omega)\}\delta(\varepsilon_f - \varepsilon_i + \omega)\, f_i(\hbar\varepsilon_i)
$$
$$
\times [1 - f_f(\hbar\varepsilon_f)], \tag{1}
$$

where the leading factor of 4 accounts for spin and valley degeneracies in graphene. This factor is the same for DLG and MIG structures, as they involve at least one graphene layer. Electrons undergo transitions from initial occupied states to final empty states of energies $\hbar\varepsilon_i$ and $\hbar\varepsilon_f$, respectively. The occupation of these levels follows the Fermi-Dirac distributions $f_{i|f}(\hbar\varepsilon_{i|f}) = 1/\{1 + \exp[(\hbar\varepsilon_{i|f} - E_{\mathrm{F},i|f})/k_{\mathrm{B}}T]\}$, where $E_{\mathrm{F},i}$ and $E_{\mathrm{F},f}$ are the corresponding chemical potentials in the emitting and receiving materials, referenced to a common origin of energies. We work here in the $T = 0$ limit, for which these distributions become step functions $f_i(\hbar\varepsilon_i) = \Theta(E_{\mathrm{F},i} - \hbar\varepsilon_i)$ and $1 - f_f(\hbar\varepsilon_f) = \Theta(\hbar\varepsilon_f - E_{\mathrm{F},f})$ and the chemical potentials reduce to the corresponding Fermi energies. For initial or final states in graphene, we recast the sum over electron states as $\sum_{i|f} \to (2\pi)^{-2}A \int d^2\mathbf{Q}_{i|f}$, while for metals it becomes $\sum_{i|f} \to (2\pi)^{-3}V \int d^3\mathbf{k}_{i|f}$, where A and V are the corresponding normalization area and volume, respectively. In the evaluation of Eq. (1), we separate the electron wave functions into parallel and perpendicular components as $\psi(\mathbf{r}) = \varphi^\parallel(\mathbf{R})\varphi^\perp(z)$, where $\mathbf{R} = (x, y)$ are in-plane coordinates. Finally, we present some results below for the loss probability normalized per unit of film surface area $J(\omega) = \Gamma(\omega)/A$ (i.e., the spectrally resolved inelastic tunneling current).

III. GRAPHENE/hBN/GRAPHENE STRUCTURES: DEPENDENCE ON DOPING LEVEL

We first discuss structures formed by two graphene layers separated by a hexagonal boron nitride (hBN) film, assuming perfect alignment between the two graphene reciprocal crystal lattices. We introduced this system in a previous publication [24], but here we study the dependence on the doping level of the two graphene layers. The tunneling structure is sketched in Fig. 1(a): a sandwich of two perfectly stacked graphene layers, with different doping levels and conductivities σ_1 (bottom) and σ_2 (top) and connected with a bias voltage V_b, are separated by a distance $d \simeq 1$ nm (about three atomic layers of hBN), which is maintained throughout this work. The first Brillouin zone of the graphene reciprocal lattice is schematically plotted in Fig. 1(b). Around the K point $[\mathbf{K} = (2\pi/3a)(1, 1/\sqrt{3})$, where $a = 1.42\text{Å}$ is the nearest-neighbor C-C distance], the electronic band structure is conical in $\mathbf{Q}$ [the two-dimensional (2D) wave vector relative to K], and we can express the electron wave functions as spinors with two components, each of them associated with one of the two carbon atoms in the unit cell [26]. This representation leads to a closed-form expression

for the graphene surface conductivity in the random-phase approximation (RPA) [27,28], which we use throughout this work with a phenomenological inelastic electron lifetime of 66 fs (10 meV energy width, or, equivalently, a moderate mean free path of 66 nm for a Fermi velocity $v_\mathrm{F} \approx 10^6$ m/s).

For the evaluation of Eq. (1), we incorporate the following elements.

Screened interaction. We use an analytical expression for the screened interaction describing the graphene/hBN/graphene sandwich, given in Appendix B, as a generalization of the results presented in Ref. [24] to an asymmetric environment. This expression incorporates the anisotropy of hBN through its parametrized dielectric function [29] for the in- and out-of-plane directions, which accounts for optical phonons in this material.

Parallel electron wave functions. We use a Dirac-fermion description of the in-plane graphene electron wave functions as

$$
\varphi_{i|f}^\parallel(\mathbf{R}) = \frac{1}{\sqrt{2A}}\, e^{i\mathbf{Q}_{i|f}\cdot\mathbf{R}} \begin{pmatrix} e^{i\phi_{i|f}/2} \\ e^{-i\phi_{i|f}/2} \end{pmatrix} e^{i\mathbf{K}\cdot\mathbf{R}}, \tag{2}
$$

where $\phi_{i|f}$ is the azimuthal angle of $\mathbf{Q}_{i|f}$ and $\mathbf{K}$ is the wave vector at the K point relative to the Γ point [see Fig. 1(b)]. We disregard intervalley scattering, so each of the two inequivalent K points in the first Brillouin zone produces an identical contribution.

Perpendicular electron wave functions. We describe the evolution of the electron along z through a one-dimensional wave function trapped in the graphene layers by potential wells, which are vertically offset due to V_b. Full details of this wave function are given elsewhere for graphene [24], while explicit expressions for metals are offered in Appendix A by treating the surface in the one-electron step-potential approximation.

The interaction between plasmons in the two neighboring layers produces a characteristic hybridization scheme, leading to two plasmon branches: optical and acoustic. Acoustic plasmons have lower energy, and their out-of-plane electric-field profile looks symmetric. In contrast, optical plasmons possess higher energy and antisymmetric field profiles, so they are generally easier to excite and manipulate and therefore are more suitable for photonic applications.

Tunneling requires $V_\mathrm{b} \neq 0$ and $E_{\mathrm{F},1} \neq E_{\mathrm{F},2}$. We take $E_{\mathrm{F},1} > E_{\mathrm{F},2}$ and hence favor tunneling from layer 1 (bottom) to layer 2 (top). We explore the effect of varying the doping-level difference between the two graphene layers by calculating the spectrally resolved inelastic tunneling current shown in Fig. 1(c), where we fix $E_{\mathrm{F},2} = 0.3$ eV and $eV_\mathrm{b} = 1.2$ eV. Energy splitting due to phonons in the hBN film can be observed at low-energy transfers (below 0.3 eV), losing strength as the Fermi energy difference $\Delta E_\mathrm{F} = E_{\mathrm{F},1} - E_{\mathrm{F},2}$ decreases.

Acoustic plasmons become apparent for $\Delta E_\mathrm{F} > 0.2$ eV. These resonances have an average plasmon lifetime of $\sim$50–70 fs [Fig. 1(f), blue dashed curve] and undergo a redshift with increasing ΔE_F [Fig. 1(f), blue solid curve]. In contrast, optical plasmons become more pronounced when $E_{\mathrm{F},1} > 2E_{\mathrm{F},2}$; they experience a milder redshift with increasing ΔE_F [Fig. 1(f), green solid curve], and their lifetime rapidly increases from $\sim$10 to $\sim$80 fs [Fig. 1(f), green dashed curve].
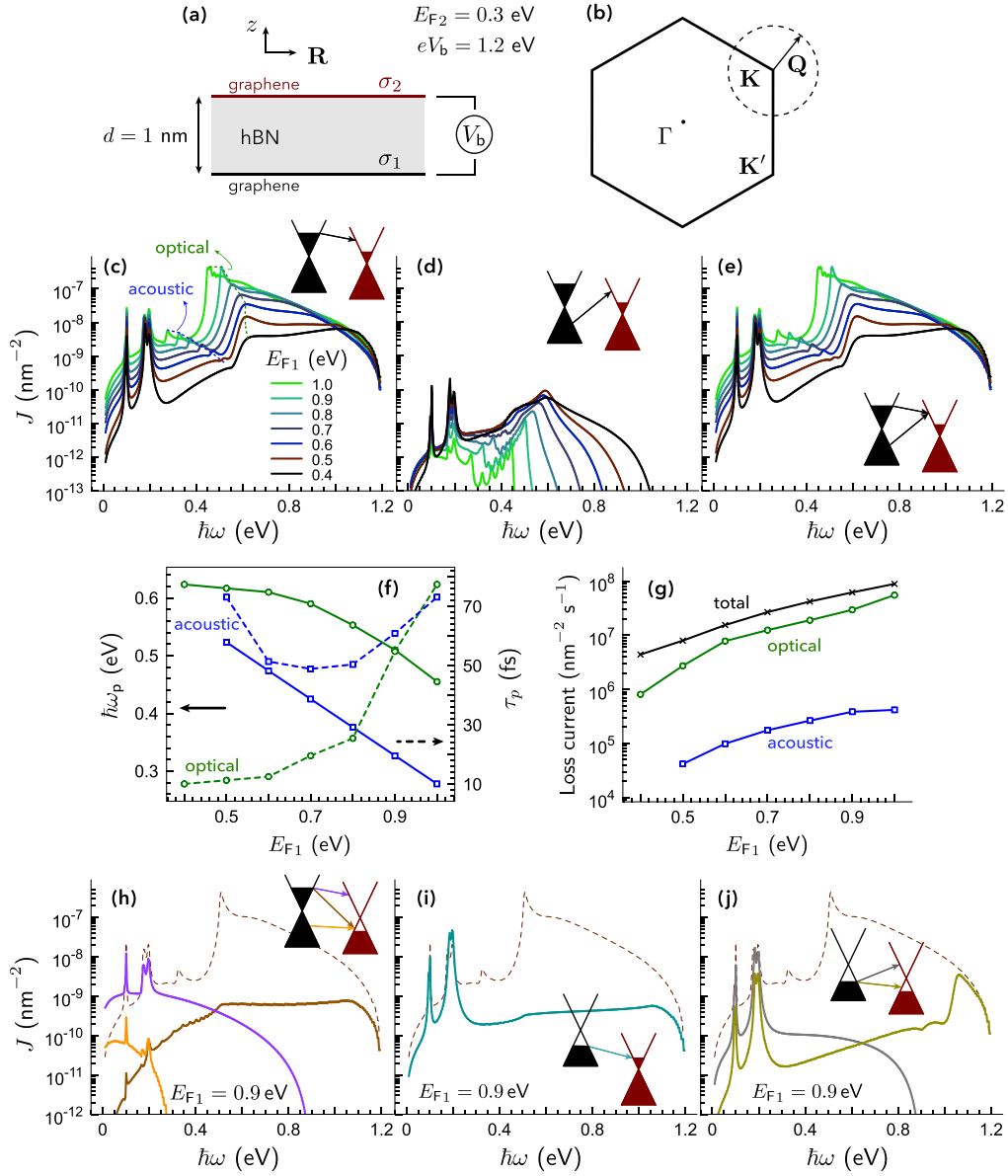
FIG. 1. Plasmon generation in double-layer graphene: dependence on doping level. (a) Asymmetrically doped double-layer graphene (DLG; upper layer at Fermi energy $E_{F,2} = 0.3$ eV, fixed in this figure, and lower layer at varying $E_{F,1}$) with an intercalated hBN film ($d = 1$ nm thickness, corresponding to about three atomic layers) and gated with a bias voltage $V_b = 1.2$ eV. (b) High-symmetry points in the graphene reciprocal lattice. (c)–(e) Spectrally resolved inelastic tunneling current for different values of $E_{F,1}$ when both graphene sheets have electron doping (see insets). We show the separate contributions of (c) conduction-to-conduction and (d) valence-to-conduction transitions, as well as (e) the sum of these two. Optical and acoustic plasmons are marked with labels in (c), while lower-energy, sharp features are associated with the excitation of hBN optical phonons. (f) Fermi energy dependence for the energies (left axis, solid curves) and lifetimes (right axis, dashed curves) of the optical (green curves) and acoustic (blue curves) plasmon modes in the DLG structure. (g) Dependence of the total inelastic tunneling current (black curve) on $E_{F,1}$, along with the partial contributions of the optical (green curve) and acoustic (blue curve) plasmons. (h)–(j) Spectrally resolved inelastic tunneling current for $E_{F,1} = 0.9$ eV with different types of doping: (h) electron-hole, (i) hole-electron, and (j) hole-hole. Possible transitions between the valence and conduction bands of the two graphene sheets are indicated by arrows in the insets. The contribution of conduction-to-conduction tunneling under electron-electron doping is shown for reference [dashed curves, taken from (c)].

We quantify the fraction of the total inelastic current invested in exciting plasmons by comparing the area under the whole spectrum in Fig. 1(g) for different values of $E_{F,1}$ [Fig. 1(g), black curve] to the area under either the optical (green curve) or acoustic (blue curve) plasmon regions. The average generation efficiency for the selected values of V_b and $E_{F,2}$ lies in the ∼20%–60% range for optical plasmons and is <1% for acoustic modes.

IV. GRAPHENE/hBN/GRAPHENE STRUCTURES: DEPENDENCE ON THE TWIST ANGLE

In practical devices, the alignment of the two graphene layers can be a challenge, so we examine the effect of a finite twist angle between their respective lattices. Rotations in real space result in rotations around the Γ point in momentum space. In Fig. 2(a) we depict two reciprocal lattices
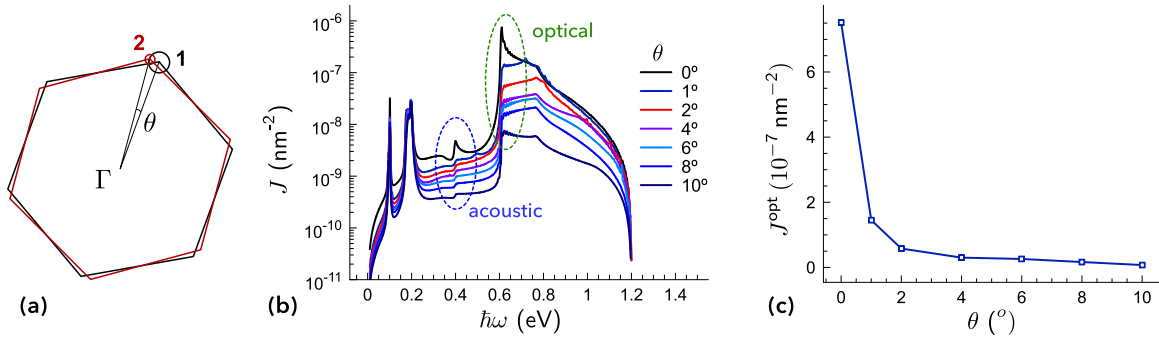
FIG. 2. Plasmon generation in double-layer graphene: dependence on tilt angle. (a) First Brillouin zones of two graphene layers tilted by an angle θ ($5°$ in this plot), along with their respective Fermi surfaces (circular cross sections of their respective Dirac cones; we plot only one per layer) for Fermi energies $E_{F,1} = 1$ eV and $E_{F,2} = 0.5$ eV. (b) Spectrally resolved inelastic tunneling current for different tilt angles and the following choice of parameters: $eV_b = 1.2$ eV, $E_{F,1} = 1$ eV, $E_{F,2} = 0.5$ eV, and $d = 1$ nm [see Fig. 1(a)]. Optical and acoustic plasmons are marked with labels, while lower-energy, sharp features are associated with the excitation of hBN optical phonons. (c) Decay of the partial contribution of optical-plasmon excitation to the inelastic current at $\hbar\omega \approx 0.61$ eV as a function of tilt angle.

corresponding to the bottom (1, black) and top (2, red) layers for a finite rotation angle θ. The circles surrounding their respective K points represent the projections of their Fermi surfaces on 2D momentum space for Fermi energies $E_{F,1} = 1$ eV and $E_{F,2} = 0.5$ eV. Inelastic tunneling requires the two circles to be either overlapping or closer to each other than the plasmon momentum, which is small compared with the Fermi momenta ($k_F = E_F/\hbar v_F$, i.e., the radii of the plotted circles), leading to a cutoff angle $\sim 10°$.

We introduce the twist angle θ in our formalism through the electron wave function of the final states, so that we maintain the initial states as in Eq. (2) with $\mathbf{K} = \mathbf{K}_i$ but write the final wave function as

$$\varphi_f^{\parallel}(\mathbf{R}) = \frac{1}{\sqrt{2A}} \, e^{i\mathbf{Q}_f \cdot \mathbf{R}} \begin{pmatrix} e^{i(\phi_f - \theta)/2} \\ e^{-i(\phi_f - \theta)/2} \end{pmatrix} e^{i\mathbf{K}_f \cdot \mathbf{R}}. \tag{3}$$

Notice that $\mathbf{K}_i$ and $\mathbf{K}_f$ are the K points corresponding to the centers of the circles in Fig. 2(a). The spinor product in Eq. (1) involving the parallel wave functions thus becomes

$$\varphi_i^{\parallel\dagger}(\mathbf{R}) \cdot \varphi_f^{\parallel}(\mathbf{R}) \varphi_f^{\parallel\dagger}(\mathbf{R}') \cdot \varphi_i^{\parallel}(\mathbf{R}')$$

$$= \frac{1}{2A^2} e^{i(\mathbf{K}_f - \mathbf{K}_i + \mathbf{Q}_f - \mathbf{Q}_i) \cdot (\mathbf{R} - \mathbf{R}')} [1 + \cos(\phi_i - \phi_f + \theta)],$$

with $\mathbf{K}_f - \mathbf{K}_i = (2\pi/3a)(\cos\theta - \sin\theta/\sqrt{3} - 1, \sin\theta + \cos\theta/\sqrt{3} - 1/\sqrt{3})$. Introducing this expression together with the parallel-wave-vector decomposition of W [Eq. (B1)] into Eq. (1) and performing the $\mathbf{R}$ and $\mathbf{R}'$ integrals, we find the condition $\mathbf{k}_{\parallel} + \mathbf{K}_f - \mathbf{K}_i + \mathbf{Q}_f - \mathbf{Q}_i = 0$, which guarantees momentum conservation. We find it convenient to separate the contributions of the perpendicular wave functions as

$$I_1(k_{\parallel}, \omega) = -\int dz \int dz' \, \varphi_i^{\perp\dagger}(z) \varphi_f^{\perp}(z) \varphi_f^{\perp\dagger}(z') \varphi_i^{\perp}(z')$$

$$\times \text{Im}\{W(k_{\parallel}, z, z', \omega)\},$$

where $\varphi_{i|f}^{\perp}$ are solutions corresponding to the two quantum wells that we use to model the graphene layers [24]. Putting these elements together and considering the initial electron to tunnel from the conduction band of layer 1, we can write from

Eq. (1) the spectrally resolved tunneling current density

$$J(\omega) = \frac{\Gamma(\omega)}{A} = \frac{e^2}{4\pi^4} \int d^2\mathbf{Q}_i \int d^2\mathbf{k}_{\parallel} \, I_1(k_{\parallel}, \omega)$$

$$\times \delta(\hbar v_F Q_f - \hbar v_F Q_i + E_{F,1} - E_{F,2} - eV_b + \hbar\omega)$$

$$\times [1 + \cos(\phi_i - \phi_f' - \theta)]\Theta(E_{F,1} - \hbar v_F Q_i)$$

$$\times \Theta(\hbar v_F Q_f - E_{F,2}), \tag{4}$$

where $\mathbf{G}_{\parallel} = \mathbf{K}_f - \mathbf{K}_i + \mathbf{k}_{\parallel}$, $\phi_f' = \tan^{-1}[(Q_{yi} - G_{\parallel_y})/(Q_{xi} - G_{\parallel_x})]$, and $Q_f = |\mathbf{Q}_i - \mathbf{G}_{\parallel}|$. We note that the Dirac δ function ensures energy conservation in Eq. (4), limiting the spectral range to $-v_F G_{\parallel} + (eV_b + E_{F,2} - E_{F,1})/\hbar < \omega < v_F G_{\parallel} + (eV_b + E_{F,2} - E_{F,1})/\hbar$. For $\theta = 0$ (no twist), this expression reduces to that of Ref. [24], where, incidentally, the Fermi energies within the delta function were absorbed in V_b and a spurious factor of 2π was accidentally introduced in the numerical results.

We represent in Fig. 2(b) the tunneling current obtained from Eq. (4) for different rotation angles when fixing the bias voltage to $V_b = 1.2$ eV and the Fermi energies to $E_{F,1} = 1$ eV and $E_{F,2} = 0.5$ eV. When plotting the maximum of $J(\omega)$ associated with the optical plasmon ($\hbar\omega_p^{opt} \simeq 0.61$ eV), we find a sharp decay with increasing twist angle [Fig. 2(c)], although we still maintain $\sim 20\%$ of the maximum value for $\theta = 4°$, further indicating a reasonable tolerance of this type of device against unintended misalignments below the $1°$ level.

So far, we have considered electron doping in both of the graphene layers. We can straightforwardly repeat the above analysis to find expressions that apply to situations in which one or both of the graphene layers is doped with holes. Furthermore, for sufficiently high V_b, electrons can tunnel from the valence band of layer 1 even when both layers have electron doping, which leads to an additional term in the integrand of Eq. (4); although we give the resulting expression below these lines, we do not consider such high voltages in this study. We find

$$J(\omega) = \frac{e^2}{4\pi^4} \int d^2\mathbf{Q}_i \int d^2\mathbf{k}_{\parallel} I_1(k_{\parallel}, \omega) \Delta(\mathbf{Q}_i, \mathbf{k}_{\parallel}, \omega),$$

where

$$
\Delta(\mathbf{Q}_i, \mathbf{k}_\parallel, \omega) = \begin{cases}
\begin{aligned}
& \delta(\hbar v_F Q_f - \hbar v_F Q_i + E_{F,1} - E_{F,2} - eV_b + \hbar\omega)C_+ \ \Theta(E_{F,1} - \hbar v_F Q_i) \ \Theta(\hbar v_F Q_f - E_{F,2}) \\
& + \delta(\hbar v_F Q_f + \hbar v_F Q_i + E_{F,1} - E_{F,2} - eV_b + \hbar\omega)C_- \ \Theta(\hbar v_F Q_f - E_{F,2}) & \text{(e-e doping)},
\end{aligned} \\[2ex]
\begin{aligned}
& \delta(-\hbar v_F Q_f - \hbar v_F Q_i + E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_- \ \Theta(E_{F,1} - \hbar v_F Q_i) \ \Theta(E_{F,2} - \hbar v_F Q_f) \\
& + \delta(\hbar v_F Q_f - \hbar v_F Q_i + E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_+ \ \Theta(E_{F,1} - \hbar v_F Q_i) \\
& + \delta(-\hbar v_F Q_f + \hbar v_F Q_i + E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_+ \ \Theta(E_{F,2} - \hbar v_F Q_f) \\
& + \delta(\hbar v_F Q_f + \hbar v_F Q_i + E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_- & \text{(e-h doping)},
\end{aligned} \\[2ex]
\begin{aligned}
& \delta(\hbar v_F Q_f + \hbar v_F Q_i - E_{F,1} - E_{F,2} - eV_b + \hbar\omega)C_- \ \Theta(\hbar v_F Q_i - E_{F,1}) \ \Theta(\hbar v_F Q_f - E_{F,2}) & \text{(h-e doping)},
\end{aligned} \\[2ex]
\begin{aligned}
& \delta(-\hbar v_F Q_f + v_F Q_i - E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_+ \ \Theta(\hbar v_F Q_i - E_{F,1}) \ \Theta(E_{F,2} - \hbar v_F Q_f) \\
& + \delta(\hbar v_F Q_f + \hbar v_F Q_i - E_{F,1} + E_{F,2} - eV_b + \hbar\omega)C_- \ \Theta(\hbar v_F Q_i - E_{F,1}) & \text{(h-h doping)},
\end{aligned}
\end{cases}
$$

with $C_\pm = 1 \pm \cos(\phi_i - \phi'_f - \theta)$. The sign in $C_\pm$ (i.e., + for valence-to-valence or conduction-to-conduction transitions, and − otherwise) originates in the change of the sign of one of the components of the spinor in Eqs. (2) and (3) needed to describe Dirac fermions in the lower Dirac cone instead of the upper one [26]. The labels e (electron) and h (hole) indicate the type of doping in the first and second layers, respectively, and several δ functions appear in each expression depending on whether the electron originates in the valence or conduction bands (for electron doping in layer 1) or just in the valence band (for hole doping in layer 1), and whether the electron tunnels to the valence or conduction bands (for hole doping in layer 2) or just to the conduction band (for electron doping in layer 2). We define $E_{F,1}$ and $E_{F,2}$ as positive quantities, although it is understood that they represent a lowering of the Fermi energy relative to the Dirac point when doping with holes.

In Figs. 1(d) and 1(h)–1(j) we show results for two perfectly stacked graphene layers with different types of doping. Arrows in the insets indicate the different tunneling channels. Additionally, we compare the results to a reference spectrally resolved tunneling probability associated with conduction-to-
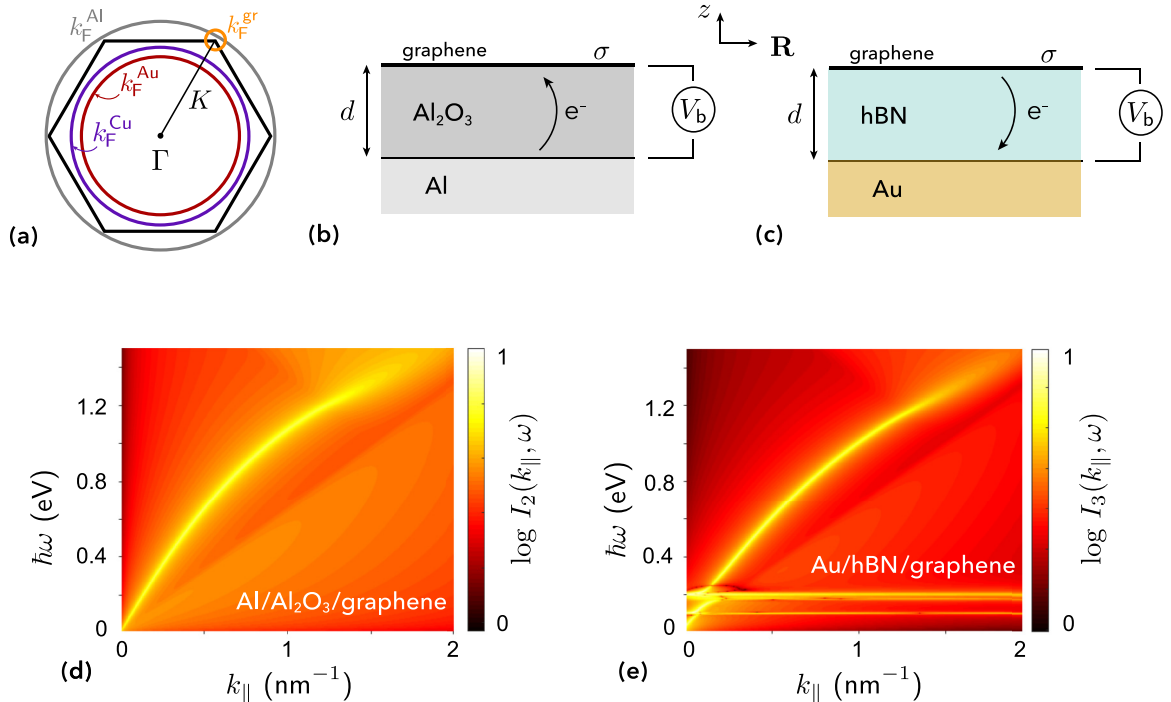

FIG. 3. Plasmon generation in metal/insulator/graphene tunneling structures. (a) Projection of the Fermi surface on the space of wave vectors parallel to the graphene plane for Al (gray), Au (red), and Cu (violet), along with the first Brillouin zone of graphene and its Fermi surface for $E_F^{gr} = 1$ eV doping (orange circle). (b) Scheme of the Al/Al$_2$O$_3$/graphene structure gated to a bias voltage V_b. The thickness of the oxide layer is $d = 1$ nm. The tunneling current goes from Al to graphene. (c) Scheme of the Au/hBN/graphene structure gated to a bias voltage V_b. The thickness of the hBN layers is also $d = 1$ nm (about three atomic layers). The tunneling current goes from graphene to Au. (d) Plasmon dispersion relation given by the poles of the screened potential for the Al/Al$_2$O$_3$/graphene structure (1 eV doping, tunneling from aluminum to graphene) when the metal is assumed to respond as a perfect conductor [see Eq. (9)]. (e) Same as (d) with Au instead of Al, hBN instead of Al$_2$O$_3$, and electrons tunneling from graphene to the metal [see Eq. (10)].

conduction transitions in two electron-doped graphene layers, which we conclude, in fact, to be the most effective configuration to generate plasmons under the conditions of the figure.

V. PLASMON GENERATION IN METAL/INSULATOR/GRAPHENE TUNNELING STRUCTURES

We now compare the performance of DLG and MIG structures. In the latter, one of the graphene layers is substituted by a semi-inifinite metal medium. When the tunneling electrons go from the metal to the graphene sheet, not all metals can comply with momentum conservation. In Fig. 3(a) we represent the surface projection of the Fermi spheres (assuming for simplicity an independent-electron description of the metal Fermi sea) of different plasmonic metals (Au, Cu, and Al) compared with the first Brillouin zone of graphene, where we further show the graphene Fermi surface for a doping level $E_F^{gr} = 1$ eV. When comparing the Fermi wave vectors of these metals ($k_F^{Au} = 12.1$ nm^{-1}, $k_F^{Ag} = 12.0$ nm^{-1}, $k_F^{Cu} = 13.6$ nm^{-1}, $k_F^{Al} = 17.5$ nm^{-1}) with the wave vector at the graphene K point ($K = 17.0$ nm^{-1}), considering the small radius of the graphene Fermi circle ($k_F^{gr} = E_F^{gr}/\hbar v_F = 1.52$ nm^{-1}), we find that only the Al Fermi sea overlaps the graphene Dirac cone, so this is the only one among the good plasmonic metals in which electrons can tunnel to graphene (without the mediation of phonons or defects). Then, it is reasonable to consider the structure represented in Fig. 3(b), consisting of a graphene layer on top of an aluminum surface coated with a 1-nm layer of oxide (Al/Al$_2$O$_3$/graphene). A gate voltage is then introduced to make metal electrons tunnel into the doped graphene sheet and excite plasmons in the MIG structure.

In contrast, when electrons tunnel from graphene to the metal, the situation is reversed, so it is favorable to have the occupied Dirac-cone region outside the projected metal Fermi sea. This situation is encountered with different choices of metal, in particular with gold, separated from graphene by three atomic layers of hBN, as depicted in Fig. 3(c) (Au/hBN/graphene).

For simplicity, we calculate the screened interaction (see Appendix B) assimilating the metal to a perfect electric conductor ($|\epsilon| \to \infty$ is a good approximation for Au and Al within the plasmonic energy range under consideration). We further approximate the response of aluminum oxide by using a constant isotropic permittivity $\epsilon_{Al_2O_3} = 3$ (the measured permittivity [30] varies by only ~4% within the energy range under consideration). Finally, we use the RPA conductivity for graphene [27,28] (see Sec. III) and the anisotropic permittivity described in Appendix B for hBN [29].

We calculate the spectrally resolved tunneling probability using Eq. (1) and the formalism described in Sec. II, with the metal wave functions (either as initial or final states, depending on the configuration) described as

$$\psi_{i|f}(\mathbf{r}) = \varphi_{i|f}^{\perp}(z, k_{i|fz}) \frac{1}{\sqrt{A}} e^{i\mathbf{k}_{i|f}^{\parallel} \cdot \mathbf{R}}.$$

The perpendicular component of these wave functions depends on the incident wave vector along z (i.e., $k_{i|fz}$) according to the explicit expressions derived in Appendix A under the assumption of a step potential to represent the metal/insulator interface.

As graphene has two sublattices, a small phase difference $e^{i\mathbf{k}_i^{\parallel} \cdot \mathbf{a}}$ has to be introduced to account for the coupling to the metal wave functions, where $\mathbf{a} = a\hat{x}$ is the C-C bond vector. Additionally, depending on the applied bias, the graphene doping level, and the type of doping (i.e., electrons or holes), only one of the two graphene Dirac cones can be engaged in the tunneling process. Taking these elements into account, the expression for the inelastic tunneling current calculated from Eq. (1) needs to be specified for each of the following MIG configurations:

(i) Al $\to$ Al$_2$O$_3$ $\to$ electron-doped graphene [see Fig. 4(a)],

$$J(\omega) = \frac{e^2}{2\pi^4} \int d^2\mathbf{Q}_f \int d^3\mathbf{k}_i \, I_2(k_{\parallel}, k_{iz}, \omega) \, \Theta\left(E_F^{Al} - \hbar^2 k_i^2/2m_e\right)[1 + \cos(\phi_f + \mathbf{k}_i^{\parallel} \cdot \boldsymbol{a})]$$
$$\times \delta\left(\hbar v_F Q_f - eV_b - E_F^{gr} - \hbar^2 k_i^2/2m_e + E_F^{Al} + \hbar\omega\right) \Theta\left(\hbar v_F Q_f - E_F^{gr}\right), \qquad (5)$$

(ii) Al $\to$ Al$_2$O$_3$ $\to$ hole-doped graphene [see Fig. 4(e)],

$$J(\omega) = \frac{e^2}{2\pi^4} \int d^2\mathbf{Q}_f \int d^3\mathbf{k}_i \, I_2(k_{\parallel}, k_{iz}, \omega) \, \Theta\left(E_F^{Al} - \hbar^2 k_i^2/2m_e\right)$$
$$\times \Big\{ \delta\left(-\hbar v_F Q_f - eV_b + E_F^{gr} - \hbar^2 k_i^2/2m_e + E_F^{Al} + \hbar\omega\right)[1 - \cos(\phi_f + \mathbf{k}_i^{\parallel} \cdot \boldsymbol{a})]\Theta\left(E_F^{gr} - \hbar v_F Q_f\right)$$
$$+ \delta\left(\hbar v_F Q_f - eV_b + E_F^{gr} - \hbar^2 k_i^2/2m_e + E_F^{Al} + \hbar\omega\right)[1 + \cos(\phi_f + \mathbf{k}_i^{\parallel} \cdot \boldsymbol{a})]\Big\}, \qquad (6)$$

(iii) electron-doped graphene $\to$ hBN $\to$ Au [see Fig. 4(f)],

$$J(\omega) = \frac{e^2}{2\pi^4} \int d^2\mathbf{Q}_i \int d^3\mathbf{k}_f \, I_3(k_{\parallel}, k_{fz}, \omega) \, \Theta\left(\hbar^2 k_f^2/2m_e - E_F^{Au}\right)$$
$$\times \Big\{ \delta\left(-\hbar v_F Q_i - eV_b + E_F^{gr} + \hbar^2 k_f^2/2m_e - E_F^{Au} + \hbar\omega\right)[1 + \cos(\phi_i + \mathbf{k}_f^{\parallel} \cdot \boldsymbol{a})] \, \Theta(E_F^{gr} - \hbar v_F Q_i)$$
$$+ \delta\left(\hbar v_F Q_i - eV_b + E_F^{gr} + \hbar^2 k_f^2/2m_e - E_F^{Au} + \hbar\omega\right)[1 - \cos(\phi_i + \mathbf{k}_f^{\parallel} \cdot \boldsymbol{a})]\Big\}, \qquad (7)$$
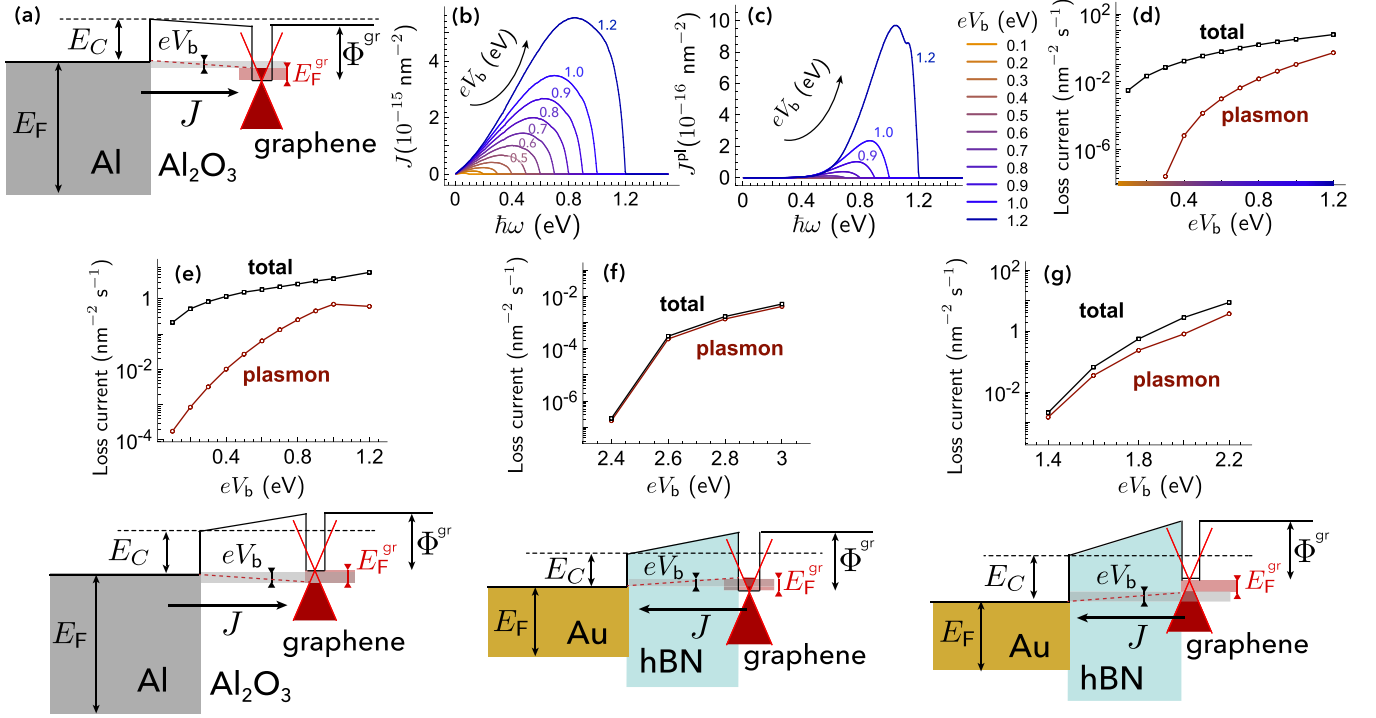
FIG. 4. Overview of plasmon generation via electron tunneling in MIG structures. (a) Energy bands to scale for the $Al/Al_2O_3/graphene$ sandwich when graphene is electron doped to $E_F^{gr} = 1$ eV. This Fermi energy is shared by the other configurations shown in this figure. (b), (c) Spectrally resolved inelastic tunneling current for different bias voltages for the system of (a): (b) total current and (c) plasmon-excitation contribution [using Eq. (11)]. The value of eV_b for each curve coincides with its spectral cutoff energy $\hbar\omega$. (d) Dependence of the total inelastic tunneling current (black) and partial plasmon contribution (red) on bias voltage after frequency integration of Eq. (5) for the system of (a). (e)–(g) Inelastic tunneling currents equivalent to those in (d) for the configurations shown in the lower insets: (e) $Al/Al_2O_3/hole$-doped graphene, (f) $Au/hBN/electron$-doped graphene, and (g) $Au/hBN/hole$-doped graphene, as obtained from the frequency integrals of Eqs. (6), (7), and (8), respectively. The graphene work function is $\Phi^{gr} = 4.7$ eV, and we choose a bias $eV_b = 0.7$ eV in the sketches.

and (iv) hole-doped graphene $\rightarrow$ hBN $\rightarrow$ Au [see Fig. 4(g)],

$$J(\omega) = \frac{e^2}{2\pi^4} \int d^2\mathbf{Q}_i \int d^3\mathbf{k}_f \, I_3(k_\parallel, k_{fz}, \omega) \, \Theta\big(\hbar^2 k_f^2/2m_e - E_F^{Au}\big)[1 - \cos(\phi_i + \mathbf{k}_f^\parallel \cdot \boldsymbol{a})]$$
$$\times \delta\big(\hbar v_F Q_i - eV_b - E_F^{gr} + \hbar^2 k_f^2/2m_e - E_F^{Au} + \hbar\omega\big) \, \Theta\big(\hbar v_F Q_i - E_F^{gr}\big), \qquad (8)$$

where the arrows indicate the direction of electron tunneling, the parallel wave vector transfer is $\mathbf{k}_\parallel = \mathbf{K} + \mathbf{Q}_{i|f} - \mathbf{k}_{f|i}^\parallel$, and we have defined

$$I_2(k_\parallel, k_{iz}, \omega) = -\int dz \int dz' \, \varphi_i^{\perp\dagger}(z, k_{iz})\varphi_f^\perp(z)\varphi_f^{\perp\dagger}(z')$$
$$\times \varphi_i^\perp(z', k_{iz}) \, \mathrm{Im}\{W(k_\parallel, z, z', \omega)\} \qquad (9)$$

for metal $\rightarrow$ graphene tunneling and

$$I_3(k_\parallel, k_{fz}, \omega) = -\int dz \int dz' \, \varphi_i^{\perp\dagger}(z)\varphi_f^\perp(z, k_{fz})\varphi_f^{\perp\dagger}(z', k_{fz})$$
$$\times \varphi_i^\perp(z') \, \mathrm{Im}\{W(k_\parallel, z, z', \omega)\} \qquad (10)$$

for graphene $\rightarrow$ metal. We note that energy conservation can be fulfilled only for energies $\hbar\omega < eV_b$ in all of these structures.

The contour plots in Figs. 3(d) and 3(e) portray Eqs. (9) and (10) for fixed values of k_{iz} and k_{fz} in the

$Al/Al_2O_3/graphene$ and $Au/hBN/graphene$ structures, respectively, with a graphene Fermi energy fixed to $E_F^{gr} = 1$ eV. They clearly reveal a plasmon mode arising from the poles of the screened potential W. Additionally, Fig. 3(e) shows low-energy features associated with hBN phonons in the $Au/hBN/graphene$ structure.

In order to quantify the amount of inelastic current associated with plasmon generation, we separate the screened interaction $W = W^{dir} + W^{ref}$ into the sum of the external (W^{dir}) and reflected (W^{ref}) components (see Appendix B). Plasmons arise from the poles of W^{ref}, originating in a denominator of the form $\eta(k_\parallel) = 1 + 4\pi i\sigma k_\parallel/\omega + \tilde{\epsilon}_2 + e^{2qd}(\tilde{\epsilon}_2 - 1 - 4\pi i\sigma k_\parallel/\omega)$, where $\tilde{\epsilon}_2 = \sqrt{\epsilon_{2x}\epsilon_{2z}}$, $q = k_\parallel\sqrt{\epsilon_{2x}/\epsilon_{2z}}$, ϵ_{2x} and ϵ_{2z} are the permittivities of the insulator (Al_2O_3 or hBN) along the in-plane (x) and out-of-plane (z) directions, and σ is the graphene conductivity. We now isolate this pole and write $W^{ref} = \eta^{-1}\tilde{W}^{ref}$. In the vicinity of the plasmon pole $k_\parallel = k_p$, we can Taylor expand $\eta(k_\parallel) =$

$\eta'(k_p)(k_\parallel - k_p)$ to first order and approximate the plasmon contribution to Eqs. (9) and (10) as

$$I_2^{pl}(k_\parallel, k_{iz}, \omega) = -\int dz \int dz' \, \varphi_i^{\perp \dagger}(z, k_{iz}) \varphi_f^\perp(z) \varphi_f^{\perp \dagger}(z')$$

$$\times \varphi_i^\perp(z', k_{iz}) \mathrm{Im} \left\{ \frac{\tilde{W}(k_\parallel, z, z', \omega)}{\eta'(k_p)(k_\parallel - k_p)} \right\}, \quad (11)$$

$$I_3^{pl}(k_\parallel, k_{fz}, \omega) = -\int dz \int dz' \, \varphi_i^{\perp \dagger}(z) \varphi_f^\perp(z, k_{fz}) \varphi_f^{\perp \dagger}(z', k_{fz})$$

$$\times \varphi_i^\perp(z') \mathrm{Im} \left\{ \frac{\tilde{W}(k_\parallel, z, z', \omega)}{\eta'(k_p)(k_\parallel - k_p)} \right\}. \quad (12)$$

By using either Eqs. (9) and (10) or Eqs. (11) and (12) in Eqs. (5)–(8), we obtain the total spectrally resolved inelastic tunneling current $J(\omega)$ or the contribution arising from plasmon generation $J^{pl}(\omega)$, respectively.

We show the calculated spectrally resolved currents $J(\omega)$ and $J^{pl}(\omega)$ in Figs. 4(b) and 4(c) for Al $\rightarrow$ Al$_2$O$_3$ $\rightarrow$ electron-doped graphene inelastic tunneling with a bias voltage in the 0.1 eV $< eV_b <$ 1.2 eV range and $E_F^{gr} = 1$ eV. A significant spectral broadening in the total current [Fig. 4(b)] originates in the availability of multiple inelastic channels associated, for example, with electron-hole pair generation in the metal. Indeed, plasmons make a relatively moderate contribution [see Fig. 4(c)], as we conclude by comparing the total to the plasmon-based ω-integrated inelastic currents [Fig. 4(d)]. The plasmon generation efficiency (number of plasmons per tunneling electron, obtained from the ratio J^{pl}/J) varies from $< 10^{-6}$ for $V_b < 0.3$ eV to 0.1 for voltages close to 1.2 eV, where the plasmon generation rate is $\sim 10^{-2}$ nm^{-2} s^{-1}. Figures 4(e)–4(g) represent the total and plasmon-based inelastic currents for the configurations depicted below the graphs, from which we conclude that Al $\rightarrow$ Al$_2$O$_3$ $\rightarrow$ hole-doped graphene exhibits better performance, with plasmon generation rates approaching ~ 1 nm^{-2} s^{-1}, while for Au/hBN/graphene structures the currents are 7 orders of magnitude smaller. Overall, comparing the performance of these MIG systems to DLG, we find the latter to be much more efficient, with inelastic currents reaching 10^8 nm^{-2} s^{-1} (see Fig. 1 and Ref. [24]).

VI. CONCLUSIONS

Our simulations reveal the suitability of DLG heterostructures as plasmon sources even under moderately unfavorable conditions produced by a finite twist angle between the two graphene sheets, with $>70\%$ efficiency for a twist angle as large as $4°$. Additionally, we find the relation $E_{F,1} = 2E_{F,2}$ between the Fermi energies of the two layers to be an optimum choice to maximize the plasmon emission rate of both acoustic and optical plasmons, although the efficiency is still in the $> 10\%$ range for order-unity variations in the Fermi energies. Furthermore, our study of MIG structures leads to efficiencies that are orders of magnitude lower than DLG.

From an intuitive viewpoint, these conclusions can be understood using the following argument. We are trying to project electrons from one electrode into the other, which requires conservation of parallel momentum, differing by just the plasmon momentum, which is small compared with the

size of the Brillouin zone. For DLG, the matching is most efficient with perfect alignment, involving large overlaps of their respective Dirac cones, while the tolerance against twisting mentioned above can be roughly quantified by the angle required to produce total mismatch between the Dirac cones, of the order of a few degrees for doping levels of 0.5–1 eV. Unfortunately, when one of the electrodes is a metal, although the conduction electron density at the surface is two to three orders of magnitude larger than that of charge carriers in graphene, they are distributed over a larger momentum-space region, rendering the overlap with the Dirac cone smaller; this effect, together with a weaker plasmon strength in MIG compared with DLG, results in much poorer plasmon generation rates.

In conclusion, DLG offers an optimum choice for the electrical generation of plasmons based upon inelastic electron tunneling, which is moderately robust against twisting of the layers (i.e., misalignment of their respective K points in the first Brillouin zone) and takes place for broad ranges of doping (i.e., it is also tolerant with respect to variations in doping).

APPENDIX A: ELECTRON WAVE FUNCTIONS IN METAL/INSULATOR INTERFACES

We approximate the electron wave functions in the metal along the z direction as those of a quantum step potential, assuming the configuration shown in the scheme in Fig. 5 (metal for $z < 0$ and insulator for $z > 0$). The Fermi energies of the metals here considered are $E_F^{Al} = 11.7$ eV and $E_F^{Au} = 5.53$ eV, referred to the bottom of the conduction band. The parameter E_C accounts for the energy jump of the insulators, for which we take $E_C^{Al_2O_3} = 3.5$ eV [31] and $E_C^{hBN} = 2.6$ eV (i.e., half of their band gap energies, under the assumption that their Fermi levels are in the center of the gap, aligned with that of the metal). Moreover, T_1 and R_1 are the transmission and reflection coefficients at the interface, respectively. By solving the time-
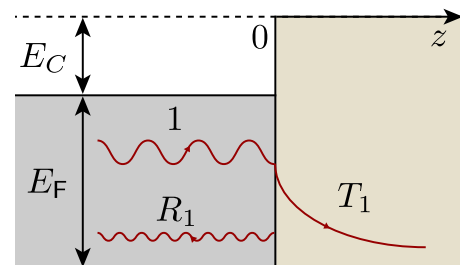

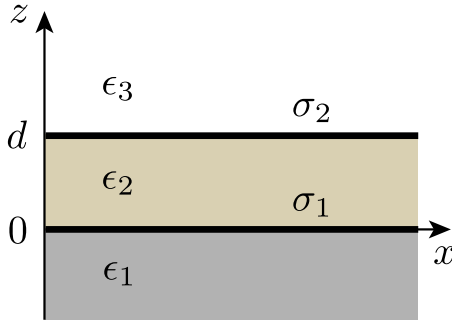FIG. 5. Configuration with metal for $z < 0$ and an insulator for $z > 0$.

FIG. 6. Multilayer structure with the permittivities $\epsilon_{1|2|3}$ and surface conductivities $\sigma_{1|2}$.

independent Schrödinger equation $(-\hbar^2/2m_e)d^2\varphi^\perp(z)/dz^2 + V(z)\varphi^\perp(z) = E\varphi^\perp(z)$ with $V(z) = -V_0\theta(-z)$ and $V_0 = E_C + E_F$, an electron incident on the interface from the metal side with energy in the insulator gap has the wave function

$$\varphi_i^\perp(z, k_{iz}) = \begin{cases} \frac{1}{\sqrt{2\pi}}(e^{ik_{iz}z} + R_1 e^{-ik_{iz}z}), & z < 0, \\ \frac{1}{\sqrt{2\pi}}T_1 e^{-k_2 z}, & z > 0, \end{cases}$$

where k_{iz} is the electron wave vector along the interface normal on the metal side, $k_2 = \sqrt{2m_e V_0/\hbar^2 - k_{iz}^2}$, $R_1 = (k_{iz} - ik_2)/(k_{iz} + ik_2)$, and $T_1 = 2k_{iz}/(k_{iz} + ik_2)$. These wave functions satisfy the orthonormality condition $\int dz\, \varphi_i^\perp(z, k_{iz})\varphi_i^{\perp\dagger}(z, k'_{iz}) = \delta(k_{iz} - k'_{iz})$. For simplicity, we assume the electron effective mass to be the same as the free electron mass.

APPENDIX B: SCREENED INTERACTION POTENTIAL

Because of the 2D translational invariance of the system, we can write the screened interaction as

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int \frac{d^2\mathbf{k}_\parallel}{(2\pi)^2} e^{i\mathbf{k}_\parallel \cdot (\mathbf{R}-\mathbf{R}')} W(k_\parallel, z, z', \omega) \quad \text{(B1)}$$

in terms of its parallel wave-vector components. We consider a multilayer structure with the permittivities $\epsilon_{1|2|3}$ and surface conductivities $\sigma_{1|2}$ as defined by the scheme in Fig. 6. This structure can describe the two configurations considered in this work: MIG (with one of the σ_j conductivities set to zero) and DLG. An expression for W was presented in a previous work [24] for a symmetric environment ($\epsilon_1 = \epsilon_3 = 1$), which we generalize here for the asymmetric configuration sketched above. Using the methods explained in that work, we separate

$$W(k_\parallel, z, z', \omega) = W^{\text{dir}}(k_\parallel, z, z', \omega) + W^{\text{ref}}(k_\parallel, z, z', \omega)$$

into direct and surface-reflection contributions, where

$$W^{\text{dir}}(k_\parallel, z, z', \omega) = \frac{2\pi}{k_\parallel} \times \begin{cases} (1/\epsilon_1)e^{-k_\parallel|z-z'|}, & z, z' < 0, \\ (1/\epsilon_3)e^{-k_\parallel|z-z'|}, & z, z' > d, \\ (1/\tilde{\epsilon}_2)e^{-q|z-z'|}, & 0 < z, z' < d, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$W^{\text{ref}}(k_\parallel, z, z', \omega) = \frac{(2\pi/k_\parallel)}{1 - A_1'A_2'e^{-2qd}} \begin{cases} (1/\epsilon_3)e^{k_\parallel(2d-z-z')}[A_2 + A_1'(A_2 + B_2')e^{-2qd}], & d < z, & d < z', \\ (1/\epsilon_3)B_2\, e^{k_\parallel(d-z)}[e^{-q(d-z')} + A_1'\, e^{-q(d+z')}], & d < z, & 0 < z' < d, \\ (\tilde{\epsilon}_2/\epsilon_1\epsilon_3)B_1 B_2\, e^{k_\parallel(d-z+z')}e^{-qd}, & d < z, & z' < 0, \\ (1/\epsilon_3)B_2\, e^{k_\parallel(d-z')}[e^{-q(d-z)} + A_1'\, e^{-q(d+z)}], & 0 < z < d, & d < z', \\ (1/\tilde{\epsilon}_2)\{A_1'\, e^{-q(z+z')} + A_2'\, e^{-q(2d-z-z')} \\ \quad + A_1'A_2'[e^{-q(2d+z-z')} + e^{-q(2d-z+z')}]\}, & 0 < z < d, & 0 < z' < d, \\ (1/\epsilon_1)B_1\, e^{k_\parallel z'}[e^{-qz} + A_2'\, e^{-q(2d-z)}], & 0 < z < d, & z' < 0, \\ (\tilde{\epsilon}_2/\epsilon_1\epsilon_3)B_1 B_2\, e^{k_\parallel(d+z-z')}e^{-qd}, & z < 0, & d < z', \\ (1/\epsilon_1)B_1\, e^{k_\parallel z}[e^{-qz'} + A_2'\, e^{-q(2d-z')}], & z < 0, & 0 < z' < d, \\ (1/\epsilon_1)e^{k_\parallel(z+z')}[A_1 + A_2'(A_1 + B_1')e^{-2qd}], & z < 0, & z' < 0. \end{cases}$$

If the material 2 is isotropic, we have $q = k_\parallel$ and $\tilde{\epsilon}_2 = \epsilon_2$. This expression is also applicable when the material in medium 2 is anisotropic (e.g., hBN) by defining $\tilde{\epsilon}_2 = \sqrt{\epsilon_{2x}\epsilon_{2z}}$ as the geometrical average of the in-plane (x) and out-of-plane (z) permittivities, with the square root chosen to yield a positive imaginary part. Also, $q = k_\parallel\sqrt{\epsilon_{2x}/\epsilon_{2z}}$ (with Re$\{q\} > 0$) is the effective out-of-plane wave vector in that medium, and we have defined the coefficients

$$B_1 = 2\epsilon_1/(\epsilon_1 + \tilde{\epsilon}_2 + \beta_1), \quad B_1' = (\tilde{\epsilon}_2/\epsilon_1)B_1, \quad B_2 = 2\epsilon_3/(\epsilon_3 + \tilde{\epsilon}_2 + \beta_2),$$
$$B_2' = (\tilde{\epsilon}_2/\epsilon_3)B_2, \quad A_j = B_j - 1, \quad A_j' = B_j' - 1, \quad \beta_j = 4\pi i k_\parallel \sigma_j/\omega$$

for $j = 1, 2$. In this work, we approximate the permittivity of alumina as a constant value $\tilde{\epsilon}_2 = \epsilon_{\text{Al}_2\text{O}_3} = 3$, and we use the perfect-conductor limit for metals ($|\epsilon_1| \to \infty$). We describe the graphene conductivity in the RPA [27,28], and we use the expression [29]

$$\epsilon_{2\ell}(\omega) \approx \epsilon_{\infty,\ell} + \sum_{i=1,2} s_{\ell i}^2 / [\omega_{\ell i}^2 - \omega(\omega + i\gamma_{\ell i})]$$

for the permittivity of hBN, where $\epsilon_{\infty,z} = 4.10$, $s_{z1} = 70.8\,\text{meV}$, $\omega_{z1} = 97.1\,\text{meV}$, $\gamma_{z1} = 0.99\,\text{meV}$, $s_{z2} = 126\,\text{meV}$, $\omega_{z2} = 187\,\text{meV}$, and $\gamma_{z2} = 9.92\,\text{meV}$ for the z component ($\ell = z$) and $\epsilon_{\infty,x} = 4.95$, $s_{x1} = 232\,\text{meV}$, $\omega_{x1} = 170\,\text{meV}$, $\gamma_{x1} = 3.6\,\text{meV}$, $s_{x2} = 43.5\,\text{meV}$, $\omega_{x2} = 95.1\,\text{meV}$, and $\gamma_{x2} = 4.3\,\text{meV}$ for the x component ($\ell = x$). The latter incorporates hBN phonons as Lorentzian terms.

[1] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, Nat. Mater. **9**, 193 (2010).

[2] N. J. Halas, S. Lal, W. Chang, S. Link, and P. Nordlander, Chem. Rev. **111**, 3913 (2011).

[3] L. M. Liz-Marzán, Langmuir **22**, 32 (2006).

[4] P. Nagpal, N. C. Lindquist, S.-H. Oh, and D. J. Norris, Science **325**, 594 (2009).

[5] S. Zeng, D. Baillargeat, H.-P. Hod, and K.-T. Yong, Chem. Soc. Rev. **43**, 3426 (2014).

[6] C. Clavero, Nat. Photonics **8**, 95 (2014).

[7] M. Danckwerts and L. Novotny, Phys. Rev. Lett. **98**, 026104 (2007).

[8] R. Zia, J. A. Schuller, A. Chandran, and M. L. Brongersma, Mater. Today **9**, 20 (2006).

[9] C. Ropers, C. C. Neacsu, T. Elsaesser, M. Albrecht, M. B. Raschke, and C. Lienau, Nano Lett. **7**, 2784 (2007).

[10] D. Alcaraz Iranzo, S. Nanot, E. J. C. Dias, I. Epstein, C. Peng, D. K. Efetov, M. B. Lundeberg, R. Parret, J. Osmond, J.-Y. Hong *et al.*, Science **360**, 291 (2018).

[11] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez *et al.*, Nature (London) **487**, 82 (2012).

[12] J. Chen, M. Badioli, P. Alonso-González, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenović, A. Centeno, A. Pesquera, P. Godignon *et al.*, Nature (London) **487**, 77 (2012).

[13] P. Bharadwaj, A. Bouhelier, and L. Novotny, Phys. Rev. Lett. **106**, 226802 (2011).

[14] A. Y. Nikitin, F. Guinea, F. J. García-Vidal, and L. Martín-Moreno, Phys. Rev. B **84**, 195446 (2011).

[15] R. H. Ritchie, Phys. Rev. **106**, 874 (1957).

[16] C. J. Powell and J. B. Swan, Phys. Rev. **115**, 869 (1959).

[17] F. J. García de Abajo, Rev. Mod. Phys. **82**, 209 (2010).

[18] B. N. J. Persson and A. Baratoff, Phys. Rev. Lett. **68**, 3224 (1992).

[19] M. Parzefall, P. Bharadwaj, A. Jain, T. Taniguchi, K. Watanabe, and L. Novotny, Nat. Nanotechnol. **10**, 1058 (2015).

[20] R. J. Walters, R. V. A. van Loon, I. Brunets, J. Schmitz, and A. Polman, Nat. Mater. **9**, 21 (2010).

[21] M. C. Zhang, T. W. Allen, and R. G. DeCorby, Appl. Phys. Lett. **103**, 071109 (2013).

[22] D. Svintsov, Zh. Devizorova, T. Otsuji, and V. Ryzhii, Phys. Rev. B **94**, 115301 (2016).

[23] V. Enaldiev, A. Btlinkin, and D. Svintsov, Phys. Rev. B **96**, 125437 (2017).

[24] S. de Vega and F. J. García de Abajo, ACS Photonics **4**, 2367 (2017).

[25] A. Woessner, A. Misra, Y. Cao, I. Torre, A. Mishchenko, M. B. Lundeberg, K. Watanabe, T. Taniguchi, M. Polini, K. S. Novoselov *et al.*, ACS Photonics **4**, 3012 (2017).

[26] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).

[27] B. Wunsch, T. Stauber, F. Sols, and F. Guinea, New J. Phys. **8**, 318 (2006).

[28] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).

[29] R. Geick, C. H. Perry, and G. Rupprecht, Phys. Rev. **146**, 543 (1966).

[30] E. D. Palik, *Handbook of Optical Constants of Solids* (Academic, San Diego, 1985).

[31] E. O. Filatova and A. S. Konashuk, J. Phys. Chem. C **119**, 20755 (2015).