Plasmons in graphene nanoribbons

F. Karimi* and I. Knezevic†

Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

(Received 30 June 2017; published 12 September 2017)

We calculate the dielectric function and plasmonic response of armchair (aGNRs) and zigzag (zGNRs) graphene nanoribbons using the self-consistent-field approach within the Markovian master equation formalism (SCF-MMEF). We accurately account for electron scattering with phonons, ionized impurities, and line-edge roughness and show that electron scattering with surface optical phonons is much more prominent in GNRs than in graphene. We calculate the loss function, plasmon dispersion, and the plasmon propagation length in supported GNRs. Midinfrared plasmons in supported (3N+2)-aGNRs can propagate as far as several microns at room temperature, with 4–5-nm-wide ribbons having the longest propagation length. In other types of aGNRs and in zGNRs, the plasmon propagation length seldom exceeds 100 nm. Plasmon propagation lengths are much longer on nonpolar (e.g., diamondlike carbon) than on polar substrates (e.g., SiO_2 or hBN), where electrons scatter strongly with surface optical phonons. We also show that the aGNR plasmon density is nearly uniform across the ribbon, while in zGNRs, because of the highly localized edge states, plasmons of different spin polarization are accumulated near the opposite edges.

DOI: 10.1103/PhysRevB.96.125417

I. INTRODUCTION

Surface plasmon polaritons or simply plasmons are collective oscillations of free carriers at the surface of good conductors [1,2]. They can be guided below the diffraction limit, i.e., their wavelength is shorter than the free-space wavelength of electromagnetic waves with the same frequency [1]. Plasmons have been studied for many applications, such as integrated nanophotonics [3–5], metamaterials [6–9], sensing [10,11], photovoltaics [12], and metasurfaces [13–15]. Plasmons were first experimentally realized at the interface of a noble metal, such as gold or silver, and a dielectric [16]. In metals, plasmons are mostly in the visible part of the electromagnetic spectrum, owing to the high carrier density [17]. Two key shortcomings of metal plasmonics are a lack of tunability of the plasmon frequency and large plasmon loss (small plasmon propagation length) [17].

Graphene plasmonics [18–26] offers several advantages over metal plasmonics. First, graphene plasmons are found in the terahertz-to-midinfrared frequency range, which is important for application in security, communications, and sensing. Second, carrier density in supported graphene is tunable by a back gate, which enables electrostatic control of graphene's electronic and optical properties [27,28]. However, the electron scattering mechanisms present in supported graphene significantly degrade the plasmon propagation length [29]. One solution is to reduce plasmon damping and improve the plasmon propagation length by lowering the system dimensionality, from two-dimensional (2D) graphene to quasi-one-dimensional (quasi-1D) graphene nanoribbons (GNRs).

Supported GNRs provide almost the same interesting plasmonic features as graphene, with the added benefit of a less dissipative environment for electrons, owing to the low electronic density of states and thus lower electron scattering rates. GNRs have different electronic and optical properties than pristine graphene and have been studied for applications such as resonators [30] and field-effect transistors [31,32]. Plasmons in *wide* GNRs have been experimentally measured [18,20,33,34]. However, in order to minimize electron scattering and obtain the best electronic and plasmonic properties, we need to look at *ultranarrow* GNRs (width < 10 nm), where very few subbands are involved in electron transport. Ultranarrow GNRs can be fabricated by top-down methods [35](conventional lithography [36], local-probe lithography [37–39], or chemical synthesis [40]) or bottom-up synthesis [41–45]. To date, there have been no experimental reports on the plasmon behavior in these structures. On the theoretical front, previous studies of plasmons in ultranarrow GNRs have relied on the random-phase approximation to treat electron-electron interaction and have either assumed ballistic (scattering-free) electron transport [46–54] or have employed the relaxation-time approximation for a simplified treatment of scattering [55,56]. However, electrons in GNRs are subject to several concurrent scattering mechanisms that are both inelastic and anisotropic and whose rates have widely varying dependencies on the electron energy; consequently, the relaxation-time approximation with a single energy-independent relaxation time is not accurate for treating dissipation in GNRs [29].

In this paper, we calculate the optical and plasmonic response of ultranarrow GNRs by employing an accurate and computationally affordable technique: the self-consistent-field approach within a Markovian master equation formalism (SCF-MMEF) for the single-electron density matrix [29]. SCF-MMEF captures interband electron-hole-pair generations and all the concurrent electron scattering mechanisms, such as with intrinsic phonons, ionized impurities, surface optical (SO) phonons, and line-edge roughness (LER). The technique uses the Bloch wave functions and dispersions calculated via the third-nearest-neighbor tight binding with a nonorthonormal basis and, in the case of zigzag GNRs, also a Hubbard term to account for the strong electron-electron interaction and spin polarization (Appendix A). We calculate the dielectric function, loss function, as well as the plasmon dispersion and propagation length in supported armchair GNRs (aGNRs)

*karimi2@wisc.edu
†iknezevic@wisc.edu

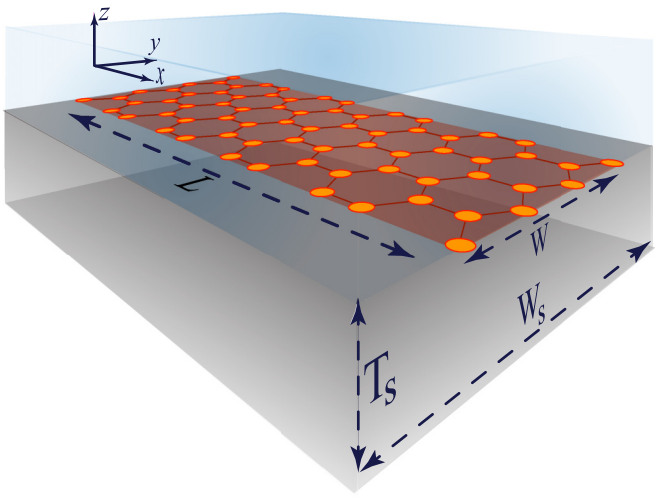

FIG. 1. Schematic of a GNR of length L and width W lying on a dielectric substrate of width W_S and thickness T_S.

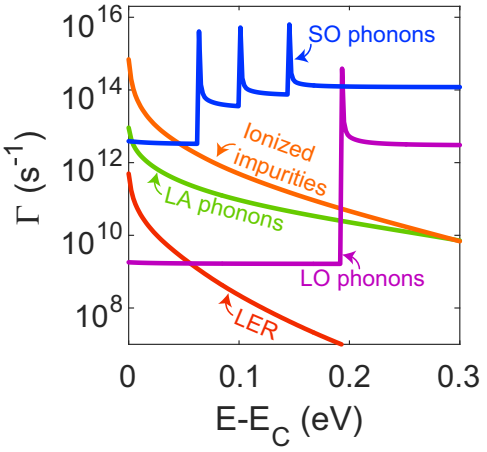

FIG. 2. Electron scattering rates for different mechanisms vs energy with respect to the top of the valence band for an n-type 35-zGNR on SiO_2. The active scattering mechanisms are longitudinal acoustic (LA) phonons, ionized impurities (II), line-edge roughness (LER), longitudinal optical (LO) phonons, and surface optical (SO) phonons from the polar substrate. The impurity density is $N_i = 4 \times 10^{11}$ cm^{-2}. The LER is assumed to be exponentially correlated, with an rms height of 0.2 Å and a correlation length of 3 nm.

and zigzag GNRs (zGNRs) on various dielectric substrates [polar SiO_2 and hBN and nonpolar diamondlike carbon (DLC)]. We show that the SO-phonon scattering rates (the interaction Hamiltonian between electrons and SO phonons in quasi-one-dimensional systems is derived in Appendix C) is at least as prominent as ionized-impurity scattering in GNRs, unlike in bulk graphene, where impurity scattering dominates (Sec. II); this has repercussions on the plasmonic response in supported GNRs. In the midinfrared, at room temperature, and for typical sheet carrier densities (up to 10^{13} cm^{-2}), plasmons in supported (3N+2)-aGNRs of width 4–5 nm can propagate as far as several microns (Sec. III). In other types of aGNRs and in zGNRs, plasmons are strongly damped and the propagation length seldom exceeds 100 nm. Plasmon propagation lengths are much longer on nonpolar (e.g., diamondlike carbon) than on polar substrates (e.g., SiO_2 or hBN), where electrons scatter strongly with surface optical phonons. Also, we obtain the plasmon distribution profile from the SCF-MMEF. aGNR plasmons are distributed almost uniformly across the ribbon (Sec. III). Owing to the highly localized edge states, zGNR plasmons are spin polarized and different spin polarizations are accumulated near the opposite edges of the ribbon (Sec. IV). We conclude with Sec. V.

II. ELECTRON SCATTERING IN GNRs SCF-MMEF OVERVIEW

We analyze a graphene nanoribbon of width W and length L, placed on a substrate of thickness T_s and width $W_S \gg W$ (Fig. 1). It is assumed that a back gate can be used to tune the carrier density and Fermi level in the GNRs and that both n-type (Fermi level in the conduction band; electrons are majority carriers) and p-type (Fermi level in the valence band; holes are majority carriers) GNRs are accessible. In order to calculate the dynamic dielectric response and plasmon properties of GNRs, we need to compute the full dynamic dielectric function, $\varepsilon(q,\omega)$, where q is the wave number in the GNR length direction (along x in Fig. 1) and ω is the angular frequency. In short, this task requires the knowledge of the electron dispersion and Bloch wave functions for the quasi-1D GNR, as well as an accurate account of all the electron dissipative pathways. In GNRs, there are several competing scattering mechanisms with different energy dependencies; some of them, like SO-phonon scattering, are both anisotropic and inelastic. Therefore, methods like the Mermin-Lindhard dielectric function [22,57] that rely on the relaxation-time approximation fail to capture the dissipation accurately. Instead, we use the SCF-MMEF, which was previously successfully applied to graphene [29], in order to calculate the dielectric function $\varepsilon(q,\omega)$ for GNRs. The SCF-MMEF provides an equation of motion for the single-particle density matrix of the electron system interacting with a perturbing electromagnetic field and a dissipative environment. The equation of motion is Markovian and guarantees that the density matrix will remain positive throughout the evolution. From it, in the frequency domain, we can obtain the information about off-diagonal terms, the so-called coherences, which are critical for calculating the polarization and from it the susceptibility and dielectric function. The dissipative terms in the SCF-MMEF equation of motion capture scattering much more rigorously than the relaxation-time approximation. Details of the calculations are given in Appendix B, as well as in Ref. [29].

Here, we will focus on building an intuitive understanding of the strength of different scattering mechanisms in GNRs. Therefore, we use Fermi's golden rule to calculate the electron scattering rates (see Appendix C) for the different mechanisms in a representative n-type armchair GNR of width approximately 4 nm on the SiO_2 substrate (Fig. 2). It is important to note that electron-SO-phonon scattering is the most prominent mechanism in GNRs at typical carrier and impurity densities; this is unlike in bulk graphene, where impurity scattering dominates. (We derive the appropriate form of the Hamiltonian for the interaction between the quasi-1D electrons in the GNR and the 2D SO phonons

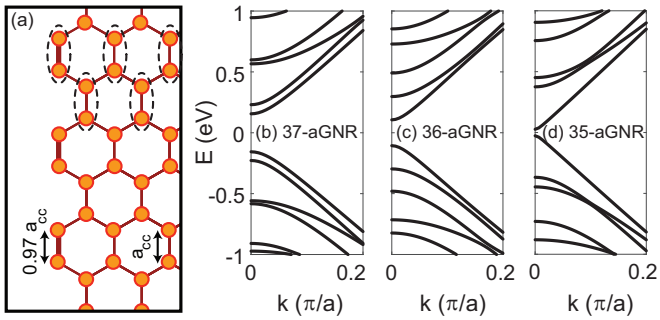

FIG. 3. (a) Schematic of an aGNR, denoting different bond lengths at the edge and near the middle. Several dimers are encircled. aGNRs are classified based on the number of dimers in a unit cell. (b)–(d) Energy dispersions of a (b) 37-aGNR, (c) 36-aGNR, and (d) 35-aGNR.

from the substrate in Appendix C 3). The relatively high SO-phonon scattering rates in narrow GNRs occur because the momentum conservation across the GNR is relaxed and each electron transition in narrow GNRs can be mediated by many SO-phonon modes. The wider the ribbon, the fewer SO phonons aid an electron transition, and, in the limit of very wide ribbons ($W \to \infty$), the momentum conservation across the ribbon is recovered and the 2D SO-phonon scattering holds. The high relative importance of electron-SO-phonon scattering means that the dielectric function and plasmonic response in narrow supported GNRs are much more sensitive to the choice of a substrate than those of bulk graphene or wide GNRs.

III. PLASMONS IN ARMCHAIR GNRs

Armchair GNRs (aGNRs) can be categorized in three groups based on the number of dimers in a unit cell: 3N, 3N+1, and 3N+2 (N is a positive integer). (3N+2)-aGNRs have the smallest band gap and are semimetallic; the rest are semiconducting. Representative band structures for the three types of aGNRs of width 4–5 nm, obtained using third-nearest-neighbor tight binding that accounts for bond shortening near the edge that accompanies hydrogen termination, are depicted in Fig. 3. Details of the band structure calculation can be found in Appendix A 1; the agreement with the results of first-principles calculation [58] is excellent.

Based on the SCF-MMEF (see Appendix B), we calculate the dynamic dielectric function $\varepsilon(q,\omega)$. In the ballistic regime (no scattering), the imaginary part of the dielectric function is zero and the plasmon dispersion is obtained from the zeros of the real part of the dielectric function [46–52]. However, in the presence of scattering, the imaginary part of the dielectric function is no longer zero. In order to obtain the plasmon dispersion, we follow the maximum of the loss function, which is proportional to $-\Im\{\frac{1}{\varepsilon(q,\omega)}\}$ and is directly measured in experiment. The loss function has a peak of finite height and width at the plasmon resonance. The higher and narrower the peak, the lower the dissipation, and the farther the plasmons can propagate. The plasmon propagation length, L_p, quantifies plasmon loss. $L_p = 1/\Delta q$, where Δq is the half width at half maximum of the loss-function peak in the wave-number direction.

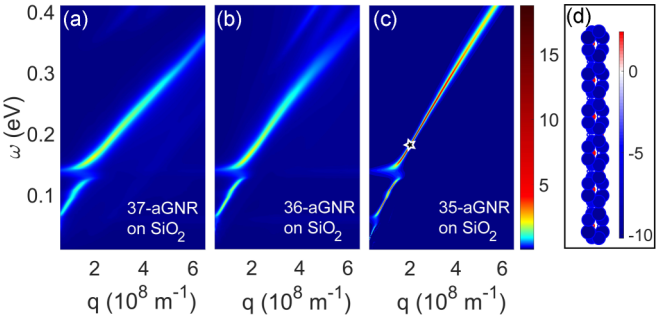

FIG. 4. (a)–(c) Loss functions in 3N+1, 3N, and 3N+2 n-type aGNRs of similar widths: (a) 37-aGNR, (b) 36-aGNR, and (c) 35-aGNR. Substrate is SiO_2, the sheet electron density is $n_s = 7 \times 10^{12}$ cm^{-2}, and the impurity density is $N_i = 4 \times 10^{11}$ cm^{-2}. (d) The plasmon spatial distribution (i.e., induced charge density distribution) at different atomic sites throughout a unit cell of the 35-aGNR at the star-marked plasmon resonance from (c). Blue is negative, red is positive, and larger circle radius means greater absolute value of the induced charge density near a site.

In Fig. 4, we show the loss functions for the same three aGNRs as in Fig. 3, which are placed on an SiO_2 substrate. All aGNRs have the same sheet electron density ($n_s = 7 \times 10^{12}$ cm^{-2}). Just by inspecting the loss functions, we expect the plasmons in (3N+2)-aGNRs to have significantly longer propagation lengths than other types of aGNRs, which we will discuss in more detail below.

Using the SCF-MMEF, we also calculate the induced charge distribution. The induced charge density distribution at the plasmon resonances corresponds to the plasmon distribution. In Fig. 4(d), we see the plasmon distribution at the star-marked plasmon resonance from Fig. 4(c). The distribution is nearly spatially uniform.

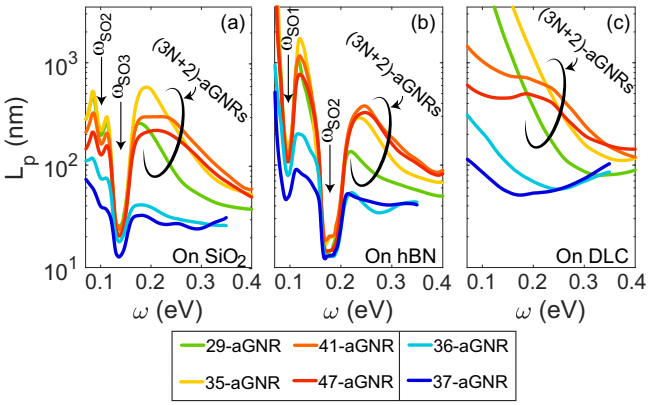

FIG. 5. Plasmon propagation length L_p for different types of a-GNRs on (a) SiO_2, (b) hBN, and (c) DLC. The electron sheet density is $n_s = 7 \times 10^{12}$ cm^{-2}. The impurity densities are $N_i = 4.0 \times 10^{11}$ cm^{-2} (SiO_2), 0.9×10^{11} cm^{-2} (hBN), and 4.2×10^{11} cm^{-2} (DLC). These impurity densities were chosen to yield the room-temperature electron mobilities of 2500 $\frac{\text{cm}^2}{\text{Vs}}$ (SiO_2), 9900 $\frac{\text{cm}^2}{\text{Vs}}$ (hBN), and 3000 $\frac{\text{cm}^2}{\text{Vs}}$ (DLC), which were reported in measurements of several groups for bulk graphene [29]. The dips in the L_p curves correspond to the hybrid plasmon-SO-phonon modes.

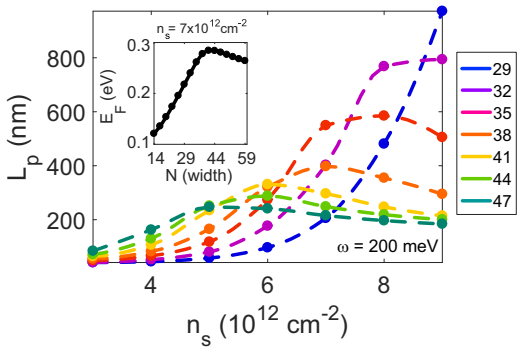

FIG. 6. Plasmon propagation length L_p vs sheet electron density for different-width (3N+2)-aGNRs on SiO_2 at the frequency $\omega = 0.2$ eV (above the highest SO-phonon angular frequency). (Inset) Fermi energy vs width for (3N+2)-aGNRs at the sheet electron density $n_s = 7 \times 10^{12}$ cm^{-2}.

Figure 5 shows the plasmon propagation length for different types of aGNRs on SiO_2, hBN, and diamond-like carbon (DLC) with the same sheet carrier density. There are two important points that can be inferred from the figure: (1) Among the three substrates, DLC provides the least dissipative environment for plasmons. As DLC is a nonpolar material, SO-phonon scattering is absent and the electron mobility is high. (2) (3N+2)-aGNRs have significantly (at least an order of magnitude) longer L_p than other aGNR types. The reason is that (3N+2)-aGNRs have the smallest band gap and lightest electron effective mass and a large separation between the first and second subband in both the conduction and valence bands, which means infrequent intersubband scattering. In contrast, (3N+1)-aGNRs not only have heavy electrons, but the first two conduction subbands are very close to one another, which results in frequent intersubband scattering and more plasmon damping (thus a shorter propagation length).

In Fig. 5, we see the effect of ribbon width. As the width increases, L_p increases, reaches a maximum, and then decreases. To get a better understanding, we plot the (3N+2)-aGNRs plasmon propagation length as a function of the sheet carrier density (n_s) for different widths at frequency $\omega = 0.2$ eV (Fig. 6). L_p has a maximum in terms of the sheet carrier density. The reason is that, by increasing the sheet carrier density, the Fermi energy rises until it reaches the second conduction subband. Thereafter, the Fermi level does not rise and more subbands get involved, so intersubband scattering rates increase. For wider ribbons, the maximum of the plasmon propagation length appears at lower sheet carrier densities, because the second conduction subband has a lower energy in wider GNRs. The inset of Fig. 6 shows the dependence of the Fermi level on the width of the ribbon for $n_s = 7 \times 10^{12}$ cm^{-2}. The maximal propagation length occurs when the Fermi level reaches the second conduction subband.

We calculated the static (dc) conductivity of (3N+2)-aGNRs on different substrates for different carrier densities (Fig. 7). We see that, for typical values of the sheet carrier density, the conductivity is highest for the 32–41-aGNRs, which corresponds to 4–5-nm width. The impurity densities are chosen so as to yield the room-temperature electron mobilities of 2500 $\frac{cm^2}{Vs}$ (SiO2), 9900 $\frac{cm^2}{Vs}$ (hBN), and 3000 $\frac{cm^2}{Vs}$ (DLC), which were reported in measurements of several groups for bulk graphene [29]. In supported bulk graphene, ionized-impurity scattering dominates and limits carrier mobility. However, in GNRs, SO-phonon scattering is very strong (Fig. 2), so the conductivity of electrons in aGNRs on the nonpolar DLC is higher than on the polar hBN, even though the bulk mobility on hBN is more than three times higher (the impurity density is nearly five times lower) than in DLC. The strong SO-phonon scattering is also the reason that peaks in the GNR conductivity on SiO_2 [Fig. 7(a)] are slightly shifted toward the narrower ribbons.

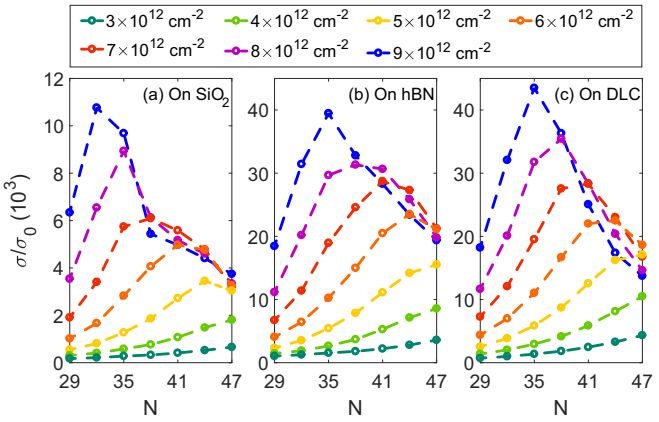

FIG. 7. Static conductivity (in the units of $\sigma_0 = e^2/h$) vs width for (3N+2)-aGNRs at different carrier densities. The impurity densities are the same as in Fig. 5: $N_i = 4.0 \times 10^{11}$ cm^{-2} (SiO_2), 0.9×10^{11} cm^{-2} (hBN), and 4.2×10^{11} cm^{-2} (DLC), chosen to yield experimentally reported room-temperature electron mobilities of 2500 $\frac{cm^2}{Vs}$ (SiO_2), 9900 $\frac{cm^2}{Vs}$ (hBN), and 3000 $\frac{cm^2}{Vs}$ (DLC). Note that the GNR conductivity on the nonpolar DLC is the same as on polar hBN, despite the bulk mobility (and thus conductivity) being over three times higher on hBN than on SiO_2.

In Fig. 8, we focus on the (3N+2)-aGNRs of different widths on DLC. The sheet electron density n_s is kept constant. Without SO-phonon scattering, we can see two clear plasmon

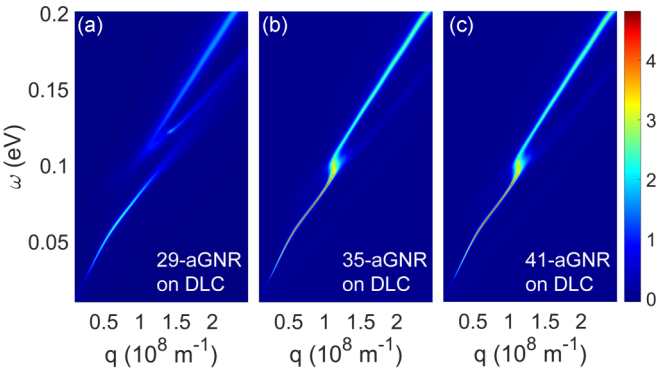

FIG. 8. (a–c) Loss functions of (3N+2) n-type aGNRs of different widths on the nonpolar DLC substrate: (a)29-aGNR, (b) 35-aGNR, and (c) 41-aGNR. The sheet electron density is $n_s = 3 \times 10^{12}$ cm^{-2} and the impurity density is $N_i = 4.2 \times 10^{11}$ cm^{-2}. Note the two plasmon branches in the ultranarrow aGNR (a), which merge into one when width is increased while keeping n_s constant. The two branches are only visible on nonpolar substrates.

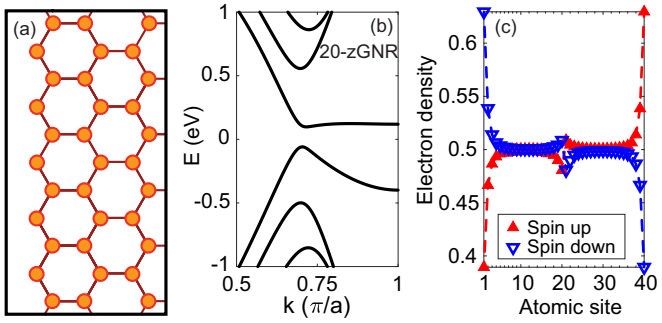

FIG. 9. (a) Schematic of a zGNR. (b) Energy dispersion for a 20-zGNR. (c) Spin distribution across the 20-zGNR. "Up" and "down" refer to two opposite spin orientations.

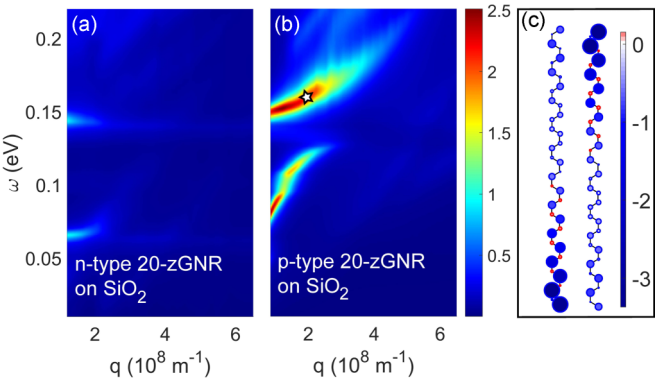

FIG. 10. Loss function of 20-zGNRs on SiO_2. (a) n-type, with a sheet electron density $n_s = 7 \times 10^{12}$ cm^{-2} and (b) p-type, with a sheet hole density $p_s = 7 \times 10^{12}$ cm^{-2}. Impurity density in (a) and (b) is $N_i = 4 \times 10^{11}$ cm^{-2}. (c) Plasmon distribution across a unit cell of the p-type 20-zGNR for two opposite spin orientations at the plasmon resonance denoted with a star in panel (b).

branches in the narrowest aGNR [Fig. 8(a)]. With increasing width, the two branches merge into one [Figs. 8(b), 8(c)]. The first branch, occurring at lower energies, emerges because of intrasubband excitations in the first conduction subband, while the second branch is related to interband excitations between the first conduction subband and the first valence subband. However, the two branches are distinct and visible only for a low Fermi level, far from the second subband; that means either low n_s or small width W (E_F depends on the line density $n_s W$). As we are keeping n_s constant and increasing W in Fig. 8, the Fermi level increases and the interband excitations get weaker (fewer empty states in the conduction band) and their rates become comparable to those of intraband excitations (both intra- and intersubband); consequently, the two branches merge.

IV. PLASMONS IN ZIGZAG GNRs

Single-electron dispersions for zGNRs are obtained using the third-nearest-neighbor tight binding (3NN TB), with an added Hubbard term in the Hamiltonian to capture the electron-electron interaction between electrons of opposite spins. The implementation details are given in Appendix A 2. Figure 9 shows the dispersions for a 20-zGNR. The 20-zGNR has a width of 4.1 nm, comparable to that of the 35-aGNR (width 4.2 nm), which has the best plasmonic response among aGNRs. Note that the electron and hole subband dispersions near the gap are fairly flat, i.e., both electron and hole masses in the lowest subband are high, with electrons being heavier ($m_e = 0.18m_0$ and $m_h = 0.08m_0$; m_0 is the rest mass of a free electron). As a result, the densities of states (DOS) in the conduction and valence subbands are also high. At typical carrier densities observed in experiment (of order 10^{12} cm^{-2}), the Fermi level is inside the gap for back-gate bias resulting in electrons as majority carriers, because electron DOS in the first subband is very high. In the case of holes as majority carriers, the Fermi level enters the valence band, as the hole DOS in the first subband is lower. Consequently, p-type zGNRs are more metallic than the n-type zGNRs, which has repercussions on the plasmonic response.

Figures 10(a) and 10(b) show the loss function for n-type and p-type 20-zGNRs, respectively, supported on SiO_2, with a sheet carrier density of 7×10^{12} cm^{-2} (the same density we analyzed for aGNRs) and the impurity density of $N_i = 4 \times 10^{11}$ cm^{-2}. In the case of the n-type zGNR, the plasmon resonance is barely visible. It is much more pronounced for the p-type zGNR, whose response is more akin to that of a metal because the Fermi level is inside the valence band. Figure 10(c) shows the plasmon spatial distribution (i.e., the induced charge density) across the zGNR for the star-marked plasmon resonance from Fig. 10(b). It is noteworthy that we have *spin-polarized edge plasmons* in p-type zGNRs at typical carrier densities. Because plasmons in zGNRs fall in the mid-infrared range and are accumulated near the edges, they could couple strongly with the vibrational modes of nearby biochemical molecules, such as proteins and DNA [59]. Therefore, p-type zGNRs might be a promising material for biosensing.

Plasmons in zGNR have short propagation lengths, as evidenced by the broad resonances in the loss function (Fig. 10); the short propagation lengths stem from the high electron and hole DOS and the resulting high scattering rates for these particles. In Fig. 11, we show the propagation

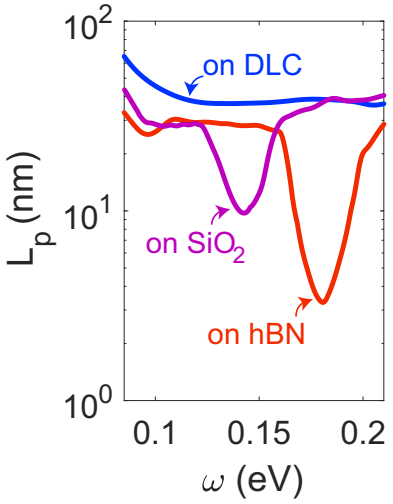

FIG. 11. Plasmon propagation length vs frequency for the p-type 20-zGNR on SiO_2, hBN, and DLC. $p_s = 7 \times 10^{12}$ cm^{-2}.

length versus frequency for the p-type 20-zGNR on different substrates. The propagation lengths are overall very short (tens of nanometers) and are lower on polar substrates (SiO_2 and hBN) than on the nonpolar DLC. The plasmon propagation length drops precipitously at the frequencies corresponding to substrate SO phonons, similar to what we observed in Fig. 5 for a-GNRs, which further underscores the importance of this scattering mechanism in the plasmonic response of GNRs.

V. CONCLUSION

In summary, we employed the SCF-MMEF along with accurate electron dispersions and Bloch wave function obtained using 3NN TB to calculate the plasmon dispersion and propagation lengths in narrow GNRs on polar (SiO_2 and hBN) and nonpolar (DLC) substrates. For narrow GNRs, SO-phonon scattering is considerably more effective than other mechanisms. For typical sheet carrier and impurity densities and at room temperature, semimetallic (3N+2)-aGNRs have the best plasmonic properties among all GNRs. Plasmons in 4–5-nm-wide (3N+2)-aGNRs can propagate as far as a few microns, with the nonpolar DLC substrate supporting the longest propagation lengths. The plasmonic response of zGNRs is weak, but zGNRs feature midinfrared spin-polarized plasmons localized near the edges, which might hold promise for the detection of biomolecules.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support by the US Department of Energy under Award No. DE-SC0008712 (Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, Physical Behavior of Materials Program). This work was performed using the compute resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences.

APPENDIX A: GNR BAND STRUCTURE

We assume that a GNR of length L (along x) and width W lies in the $z = d$ plane, as shown in Fig. 1. A substrate of dielectric constant $\varepsilon_s(\omega) = \varepsilon_0\kappa_s(\omega)$ fills the $z < 0$ half-space. The substrate volume is $V_s = LW_sT_s$, where L, W_s, and T_s are its length, width, and thickness, respectively.

To calculate the energy dispersion and wave functions, we use the tight-binding Hamiltonian

$$\mathbb{H}_e = \mathbb{H}_{TB} \equiv \sum_{\boldsymbol{R}_i^n,\boldsymbol{R}_j^m,\nu} \left(t_{\boldsymbol{R}_i^n,\boldsymbol{R}_j^m} c^\dagger_{\boldsymbol{R}_i^n\nu} c_{\boldsymbol{R}_j^m\nu} + \text{H.c.}\right), \tag{A1}$$

where $c^\dagger_{\boldsymbol{R}_i^n\nu}$ and $c_{\boldsymbol{R}_i^n\nu}$ are electron creation and destruction operators at atomic site $\boldsymbol{R}_i^n$ with spin ν, respectively. $\boldsymbol{R}_i^n = \boldsymbol{X}_n + \boldsymbol{r}_i$ with $\boldsymbol{X}_n$ being the position vector of the nth unit cell, and $\boldsymbol{r}_i$ being the position vector of the ith atomic site in the unit cell. The eigenkets and eigenenergies of $\mathbb{H}_e$ are represented as $|kl\nu\rangle$ and $\epsilon_{kl\nu}$, respectively. k is the wave vector along the x axis, l is the band index, and ν denotes the spin orientation. The TB wave function reads

$$|kl\nu\rangle = \frac{1}{\sqrt{N_{uc}}} \sum_{\boldsymbol{R}_j^m\nu} e^{ik\hat{x}\cdot\boldsymbol{R}_j^m} \mathcal{C}_{kl\nu,j} c^\dagger_{\boldsymbol{R}_j^m\nu} |\varnothing\rangle, \tag{A2}$$

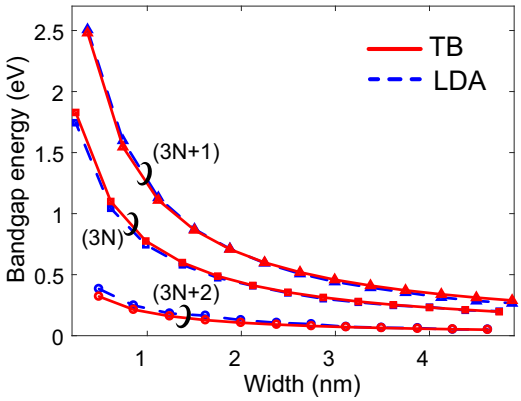

FIG. 12. Band gap versus width for 3N, 3N+1, and 3N+2 aGNRs as obtained from first principles within the LDA approximation [58] (dashed line) and from our tight-binding calculation (solid line).

where $|\varnothing\rangle$ is the vacuum state. To obtain $|kl\nu\rangle$ and $\epsilon_{kl\nu}$, we write the secular equations derived from the Schrödinger equation

$$\langle\varnothing|c_{\boldsymbol{r}_i\nu}\mathbb{H}_e|kl\nu\rangle = \epsilon_{kl\nu}\langle\varnothing|c_{\boldsymbol{r}_i\nu}|kl\nu\rangle, \forall \boldsymbol{r}_i, \tag{A3}$$

and solve them as a generalized eigenvalue problem. However, to do so, the hopping energies $\gamma_{\boldsymbol{R}_i^n\nu,\boldsymbol{R}_j^m\eta} = \langle\varnothing|c_{\boldsymbol{R}_i^n\nu}\mathbb{H}_{TB}c^\dagger_{\boldsymbol{R}_j^m\nu}|\varnothing\rangle\delta_{\nu,\eta}$ and overlap elements $\langle\varnothing|c_{\boldsymbol{R}_i^n\nu}c^\dagger_{\boldsymbol{R}_j^m\nu}|\varnothing\rangle$ are required. It should be noted that in general, $c^\dagger_{\boldsymbol{R}_i^n\nu}|\varnothing\rangle$ does not generate an orthonormal basis. So, we define the overlap matrix $\mathcal{S}$ as

$$\mathcal{S}_{\boldsymbol{R}_i^n\nu,\boldsymbol{R}_j^m\eta} = \langle\varnothing|c_{\boldsymbol{R}_i^n\nu}c^\dagger_{\boldsymbol{R}_j^m\nu}|\varnothing\rangle\delta_{\nu,\eta}. \tag{A4}$$

1. aGNRs

In aGNRs, we drop the spin quantum number, as the solutions are spin degenerate. The values we used for γ and $\mathcal{S}$ are from Ref. [60] and benchmarked against first-principles calculations in Ref. [58]:

neighbor	γ(eV)	$\mathcal{S}$
1^{st}	2.78	0.117
2^{nd}	0.15	0.07
3^{rd}	0.095	0.023

(A5)

Because there is no translational symmetry in the y direction, the bonds lengths are not all the same (see Fig. 3). The bond lengths at the edges of the hydrogen-passivated aGNRs are assumed to be $0.97a_{cc}$, $a_{cc} = 1.42$ Å being the bulk carbon-carbon bond length (Fig. 3). To calculate the overlap integrals of $(kl|k+ql')$, we define $\tilde{\mathcal{S}}$ as

$$\tilde{\mathcal{S}}_{j\nu,i\eta} = \sum_m \mathcal{S}_{\boldsymbol{R}_j^m\nu,\boldsymbol{r}_i\eta} e^{i\frac{k+k'}{2}\hat{x}\cdot(\boldsymbol{r}_i-\boldsymbol{R}_j^m)}, \tag{A6}$$

then the overlap integrals are

$$(kl\eta|k+q,l'\eta) = \tilde{\mathcal{C}}^\dagger_{kl\eta}\tilde{\mathcal{S}}\tilde{\mathcal{C}}_{kl\eta}, \tag{A7}$$

where $\tilde{\mathcal{C}}$ vector is defined as $\tilde{\mathcal{C}}_{kl\eta,i} = \mathcal{C}_{kl\eta,i}$. In Fig. 12, we see that there is excellent agreement between the tight-binding calculation and the density functional theory (DFT) calculation within the local-density approximation (LDA).

2. zGNRs

In order to capture the strong electron-electron interaction in the zGNRs, we amend $\mathbb{H}_{\rm TB}$ with the Hubbard Hamiltonian within the mean-field approximation

$$\mathbb{H}_{\rm e} = \mathbb{H}_{\rm TB} + U \sum_{\boldsymbol{R}_i^n, \nu} c^\dagger_{\boldsymbol{R}_i^n \nu} c_{\boldsymbol{R}_i^n \nu} \langle c^\dagger_{\boldsymbol{R}_i^n \bar{\nu}} c_{\boldsymbol{R}_i^n \bar{\nu}} \rangle, \tag{A8}$$

where U is the Hubbard factor, $\bar{\nu}$ is the opposite of spin ν, and $\langle\cdot\rangle$ denotes the zero-temperature expectation value corresponding to the many-particle ground state

$$|GS\rangle = \left(\prod_{kl\nu} \hat{\psi}^\dagger_{kl\nu} \right) |\varnothing\rangle, \tag{A9}$$

where $\hat{\psi}^\dagger_{kl\nu}$ creates an electron at state $|kl\nu\rangle$. In a nonorthonormal basis, $c^\dagger_{\boldsymbol{R}_i^n \nu} c_{\boldsymbol{R}_i^n \nu}$ and the number operator $n_{\boldsymbol{R}_i^n \nu}$ are not necessarily equal. The overlap matrix $\mathcal{S}$ is Hermitian and positive definite, so it can be decomposed as $\mathcal{S} = \mathcal{U}\mathcal{D}\mathcal{U}^\dagger$, where $\mathcal{U}$ is a unitary matrix, and $\mathcal{D}$ is a positive-definite diagonal matrix. As a result, $\mathcal{S}^{\frac{1}{2}}$, and $\mathcal{S}^{-\frac{1}{2}}$ are well defined. The orthonormal basis of $C^\dagger$ operators may be obtained through the so-called Lowdin transformation,

$$C^\dagger_{\boldsymbol{R}_j^m \nu} = \sum_i (\mathcal{S}^{-\frac{1}{2}})_{\boldsymbol{R}_i^n \nu, \boldsymbol{R}_j^m \nu} c^\dagger_{\boldsymbol{R}_i^n \nu}. \tag{A10}$$

One can prove $\{c_{\boldsymbol{R}\nu}, c^\dagger_{\boldsymbol{R}'\nu'}\} = \mathcal{S}_{\boldsymbol{R}\nu, \boldsymbol{R}'\nu'}$. Knowing the anticommutation relation, we obtain

$$\langle c^\dagger_{\boldsymbol{R}_i^n \nu} c_{\boldsymbol{R}_i^n \nu} \rangle = \sum_{kl} \left| \sum_{\boldsymbol{R}_j^m \eta} \frac{e^{ik\hat{x}\cdot(\boldsymbol{R}_j^m - \boldsymbol{R}_i^n)}}{\sqrt{N_{\rm uc}}} \mathcal{S}_{\boldsymbol{R}_i^n \nu, \boldsymbol{R}_j^m \eta} \mathcal{C}_{kl\eta,m} \right|^2. \tag{A11}$$

The number operator is $n_{\boldsymbol{R}_i^n \nu} = C^\dagger_{\boldsymbol{R}_i^n \nu} C_{\boldsymbol{R}_i^n \nu}$. So, the expectation value of the number of electrons at a specific site is

$$\langle n_{\boldsymbol{R}_i^n \nu} \rangle = \sum_{kl} \left| \sum_{\boldsymbol{R}_j^m \eta} \frac{e^{ik\hat{x}\cdot(\boldsymbol{R}_j^m - \boldsymbol{R}_i^n)}}{\sqrt{N_{uc}}} (\mathcal{S}^{\frac{1}{2}})_{\boldsymbol{R}_i^n \nu, \boldsymbol{R}_j^m \eta} \mathcal{C}_{kl\eta,m} \right|^2. \tag{A12}$$

In our calculations for zGNRs, $U = -2.2$ eV, while the γ and $\mathcal{S}$ values are from Ref. [61]:

neighbor	γ(eV)	$\mathcal{S}$
1^{st}	2.78	0.117
2^{nd}	0.09	0.045
3^{rd}	0.27	0.065

(A13)

We neglect the off-diagonal terms due to the Hubbard term. For calculating the overlap integrals of $(kl\nu|k + ql'\nu)$, we use equations (A6) and (A7). In Fig. 13, we see the excellent agreement between the tight-binding calculation and the density functional theory (DFT) calculation within the local-density approximation (LDA) [58].

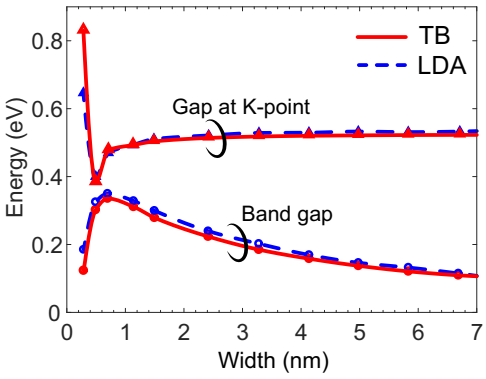

FIG. 13. Band gap and the gap at K point versus width for zGNRs as obtained from first principles within the LDA approximation [58] (dashed line) and from our tight-binding calculation (solid line).

APPENDIX B: SCF-MMEF IN 1D

In Ref. [29], we derived the SCF-MMEF for a quasi-two-dimensional system. Here, we re-derive the SCF-MMEF for a quasi-one-dimensional system. It should be noted that the SCF-MMEF can be generalized for nanoislands and nanodisks, which are essentially large molecules and thus quasi-zero-dimensional.

The induced charge density in GNRs can be written as

$$n_{\rm ind}(x,y,z,t) = n_l(x,t) \left[\frac{1}{W} \Pi\left(\frac{y}{W} \right) \right] \delta(z). \tag{B1}$$

The inhomogeneous electromagnetic wave equation describing the induced potential energy is

$$\left[\frac{\partial}{\partial^2 z} + \frac{\partial}{\partial^2 y} + (i\,Q)^2 \right] V_{\rm ind}(q,y,z,\omega) = -\frac{e^2}{\varepsilon_r \varepsilon_0} n_l(q,\omega) \left[\frac{1}{W} \Pi\left(\frac{y}{W} \right) \right] \delta(z), \tag{B2}$$

where we took the temporal Fourier transform and the spatial Fourier transform over the x coordinate. Also, $(i\,Q)^2 = \frac{\varepsilon_r \omega^2}{c^2} - \boldsymbol{q}\cdot\boldsymbol{q}$. We use the standard Green's function analysis to solve Eq. (B2):

$$\overline{V}_{\rm ind}(q, z=0, \omega) = \frac{-e}{2\pi \varepsilon_r \varepsilon_0} n_l(q,\omega) \mathfrak{I}(q), \tag{B3}$$

where the overbar denotes the average along the width of the system, i.e., $\overline{V} = \int_{-\frac{W}{2}}^{\frac{W}{2}} V(y)$. To simplify the notation, henceforth we drop the overbar and $z = 0$ but keep in mind that the potential energies are at averaged over y at $z = 0$. Also, in Eq. (B3) $\mathcal{I}(q) = \int_{\frac{-1}{2}}^{\frac{1}{2}} \int_{\frac{-1}{2}}^{\frac{1}{2}} d\eta' d\eta K_0(W|q(\eta - \eta')|)$, where K_0 is the modified Bessel function of the second kind. We define the linear polarization as

$$P_l(q,\omega) = \frac{-e n_l(q,\omega)}{V_{\rm SCF}(q,\omega)}, \tag{B4}$$

where $V_{\rm SCF}$ is the self-consistent field, i.e., the sum of external and induced fields. The susceptibility reads

$$\chi(q,\omega) \equiv \frac{V_{\rm ind}(q,\omega)}{V_{\rm SCF}(q,\omega)} = \frac{-e}{2\pi \varepsilon_r \varepsilon_0} \mathcal{I}(q) P_l(q,\omega), \tag{B5}$$

and consequently the dielectric function is

$$\varepsilon(q,\omega) = 1 + \frac{e}{2\pi\varepsilon_r\varepsilon_0}\mathcal{I}(q)P_l(q,\omega). \tag{B6}$$

Also, by solving the continuity equation, we obtain the conductivity

$$\sigma(q,\omega) = \frac{-ie\omega W}{q^2}P_l(q,\omega). \tag{B7}$$

To obtain the dielectric function [Eq. (B6)] and conductivity [Eq. (B7)], we need to know the linear polarization. We obtain P_l quantum mechanically. The induced charge density in the second quantization representation is

$$n(\boldsymbol{r},\omega) = -\frac{e}{N_{uc}}\sum_{k,q,l',l} u^*_{k+ql'}(\boldsymbol{r})u_{kl}(\boldsymbol{r})e^{-iq\cdot r}\langle c^\dagger_{k+ql'}c_{kl}\rangle(\omega). \tag{B8}$$

So, the linear polarization in the second quantization representation reads

$$P_l(q,\omega) = \frac{-e}{L}\sum_{k,l',l}\frac{\langle c^\dagger_{kl}c_{k+ql'}\rangle}{V_{\mathrm{SCF}}(q,\omega)}(kl|k+ql'). \tag{B9}$$

To calculate $\frac{\langle c^\dagger_{kl}c_{k+ql'}\rangle}{V_{\mathrm{SCF}}(q,\omega)}$, we exploit the SCF-MMEF [29]:

$$\begin{aligned}(\epsilon_{kl}-\epsilon_{k+ql'}+\hbar\omega)\langle c^\dagger_{kl}c_{k+ql'}\rangle &= (f_{kl}-f_{k+ql'})(k+ql'|kl)V_{\mathrm{eff}}(q,\omega)\\
&+i\hbar\sum_{\substack{k'mm'v\\ \epsilon_{k'm}=\epsilon_{kl}\mp\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{kl}}{\left|\frac{\partial\epsilon_{k',m}}{\partial k'}\right|}(k+ql'|k'+qm')(k'm|kl)\langle c^\dagger_{k'm}c_{k'+qm'}\rangle\\
&+i\hbar\sum_{\substack{k'mm'v\\ \epsilon_{k'+qm'}=\epsilon_{k+ql'}\mp\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k+ql'}}{\left|\frac{\partial\epsilon_{k'+q,m'}}{\partial k'}\right|}(k+ql'|k'+qm')(k'm|kl)\langle c^\dagger_{k'm}c_{k'+qm'}\rangle\\
&-i\hbar\sum_{\substack{k'mm'v\\ \epsilon_{k'm'}=\epsilon_{km}\pm\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k'm'}}{\left|\frac{\partial\epsilon_{k',m'}}{\partial k'}\right|}(km|k'm')(k'm'|kl)\langle c^\dagger_{km}c_{k+ql'}\rangle\\
&-i\hbar\sum_{\substack{k'mm'v\\ \epsilon_{k'+qm}=\epsilon_{k+qm'}\pm\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k'+qm}}{\left|\frac{\partial\epsilon_{k'+q,m}}{\partial k'}\right|}(k+ql'|k'+qm)(k'+qm|k+qm')\langle c^\dagger_{kl}c_{k+qm'}\rangle.\end{aligned} \tag{B10}$$

By discretizing the Brillouin zone, Eq. (B10) can be written in the matrix form and solved numerically for $\mathcal{X}$.

$$\mathcal{E}\mathcal{X} = \mathcal{F} + i\hbar(\mathcal{R}-\mathcal{R}'-\mathcal{R}'')\mathcal{X}. \tag{B11}$$

Each pair (q,ω) results in its own Eq. (B11). The matrices and vectors in Eq. (B11) are defined as

$$\mathcal{E}_{\{kl'l\}\{k'm'm\}} = \delta_{\{kl'l\}\{k'm'm\}}(\epsilon_{kl}-\epsilon_{k+ql'}+\hbar\omega), \tag{B12a}$$

$$\mathcal{X}_{\{kl'l\}} = \langle c^\dagger_{kl}c_{k+ql'}\rangle, \tag{B12b}$$

$$\mathcal{F}_{\{kl'l\}} = (f_{kl}-f_{k+ql'})(k+ql'|kl), \tag{B12c}$$

$$\begin{aligned}\mathcal{R}_{\{kl'l\}\{k'm'm\}} &= \sum_{\substack{v\\ \epsilon_{k'm}=\epsilon_{kl}\mp\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{kl}}{\left|\frac{\partial\epsilon_{k',m}}{\partial k'}\right|}(k+ql'|k'+qm')(k'm|kl)\\
&+\sum_{\substack{v\\ \epsilon_{k'+qm'}=\epsilon_{k+ql'}\mp\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k+ql'}}{\left|\frac{\partial\epsilon_{k'+q,m'}}{\partial k'}\right|}(k+ql'|k'+qm')(k'm|kl),\end{aligned}$$

$$\mathcal{R}'_{\{kl'l\}\{km'l\}} = \sum_{\substack{k'mv\\ \epsilon_{k'+qm}=\epsilon_{k+qm'}\pm\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k'+qm}}{\left|\frac{\partial\epsilon_{k'+q,m}}{\partial k'}\right|}(k+ql'|k'+qm)(k'+qm|k+qm'),$$

$$\mathcal{R}''_{\{kl'l\}\{kl'm\}} = \sum_{\substack{k'm'v\\ \epsilon_{k'm'}=\epsilon_{km}\pm\hbar\omega_v}}\frac{\mathcal{W}^{\pm}_{k'-k,v}\pm\Delta\mathcal{W}_{k'-k,v}f_{k'm'}}{\left|\frac{\partial\epsilon_{k',m'}}{\partial k'}\right|}(km|k'm')(k'm'|kl). \tag{B12d}$$

The set of variables $\{kl'l\}$ corresponds to a position in the $\mathcal{X}$ vector. Equation (B11) is solved for $\mathcal{X}$ for every (q,ω). We introduce a vector $\mathcal{V}$ as $\mathcal{V}_{\{kl'l\}} = (kl|k+ql')$, so the polarization is

$$P_l(q,\omega) = \frac{-2e}{L}\mathcal{V}^T\mathcal{X}. \tag{B13}$$

The factor of 2 captures the spin degeneracy, like in aGNRs. In the case of zGNRs, where the spin degeneracy is broken, Eq. (B10) should be solved for each spin orientation separately, and the total linear polarization is the sum of the linear polarization corresponding to each spin. By knowing P_l, we calculate the dielectric function [Eq. (B6)] and the conductivity [Eq. (B7)].

APPENDIX C: INTERACTION HAMILTONIAN IN QUASI-1D SYSTEMS

As before, we assume that a GNR of length L (along x) and width W lies in the $z = d$ plane, as shown in Fig. 1. A substrate of dielectric constant $\varepsilon_s(\omega) = \varepsilon_0\kappa_s(\omega)$ fills the $z < 0$ half-space. The substrate volume is $V_s = LW_sT_s$, where L, W_s, and T_s are its length, width, and thickness, respectively. For calculation of the electron energy dispersion and the Bloch wave functions, see Appendix A.

For each perturbation operator $[\delta\mathbb{V}(\boldsymbol{r})]$, we define $\mathcal{M}(q) = \overline{\delta V}(q)$ as the interaction strength. $\delta\overline{V}(q)$ is the Fourier transfer of $\delta\mathbb{V}(\boldsymbol{r})$ over x and the overline denotes the average over y. All the interaction Hamiltonians can be written in the following general form

$$\mathbb{H}_{\text{int}} = \sum_{kq,l'l} \mathcal{M}_{\text{int}}(q)(k+ql'|kl)c^\dagger_{k+ql'}c_{kl}\otimes\mathbb{B}_{q,v}, \tag{C1}$$

where c and $c^\dagger$ are the electron annihilation and creation operators, respectively. For phononic mechanisms $\mathbb{B}_{q,v} = b_{q,v} + b^\dagger_{-q,v}$, while b and $b^\dagger$ are the phonon annihilation and creation operators, respectively. The overlap integrals are defined as $(k+ql'|kl) = \int_{uc} d^3r u^*_{k+ql'}(\boldsymbol{r})u_{kl}(\boldsymbol{r})$. The details of their calculation is in Appendix A. For nonphononic mechanisms, $\mathbb{B}_{q,v}$ is an identity operator. Based on Fermi's golden rule, the scattering rate $[\Gamma^\pm_v(kl)]$ for the mechanism v is

$$\Gamma^\pm_v(kl) = \sum_{\substack{l' \\ \epsilon_{k'l'}=\epsilon_{kl}\pm\hbar\omega_v}} 2\left|\frac{\partial\epsilon_{k'l'}}{\partial k'}\right|^{-1}\mathcal{W}^\pm_{k'-k,v}|(k'l'|kl)|^2, \tag{C2}$$

where $\mathcal{W}^\pm_{k'-k,v}$ is the scattering weight. For the phononic mechanisms, $\mathcal{W}^\pm_{k'-k,v} = \frac{L}{2\hbar}|\mathcal{M}_{\text{int}}(k'-k)|^2(N_v+\frac{1}{2}\pm\frac{1}{2})$, N_v being the number of vth-type phonons. $+$ and $-$ signs correspond to absorption and emission processes, respectively. For the nonphononic mechanisms, $\mathcal{W}^\pm_{k'-k,v} = \frac{L}{4\hbar}|\mathcal{M}_{\text{int}}(k'-k)|^2$.

1. Electron scattering with intrinsic phonons

The scattering weight for the acoustic (LA) and nonpolar optical (LO) phonons can be readily written as

$$\mathcal{W}^\pm_{k'-k,\text{LA}} = \frac{1}{2}\frac{D^2_{LA}k_BT}{2\hbar\varrho v_s^2},$$
$$\mathcal{W}^\pm_{k'-k,\text{LO}} = \frac{1}{2}\left(N_{\text{LO}}+\frac{1}{2}\pm\frac{1}{2}\right)\frac{D^2_{LO}}{2\varrho\omega_{\text{LO}}}, \tag{C3}$$

where ϱ is the line mass density of the GNR. We approximate $\omega_{q,\text{LA}} = v_s|q|$ where $v_s = 2\times10^4\frac{\text{m}}{\text{s}}$, and $\omega_{q,\text{LO}} = \omega_{\text{LO}} = 194$ meV. $D_{\text{LA}} = 12$ meV and $D_{\text{LO}} = 50\frac{\text{eV}}{\text{nm}}$ are the deformation potentials. We assume the LA phonon scattering elastic.

2. Electron scattering with ionized impurities

The bare potential energy seen at $z = d$ due to an impurity located at $\boldsymbol{R}_i = (x_i,y_i,z_i)$ in the substrate is

$$\delta V(q,y)|_{\boldsymbol{R}_i} = \frac{-e^2K_0(|q\sqrt{(y-y_i)^2+(d-z_i)^2}|)}{2\pi\varepsilon_0\kappa_bL}e^{-iqx_i}, \tag{C4}$$

where we took the Fourier transform with respect to x. $\kappa_b = \frac{1+\kappa_s(\omega=0)}{2}$ denotes the background permittivity and $K_0(\cdot)$ is the second type modified Bessel function. To find the total electric potential energy, we should take a sum of Eq. (C4) over the location of all impurities. But, we only know the average density of impurities. So, we take an average of $\Phi(q)$ over all possible ensemble of impurities, ($\{\boldsymbol{R}_i\}$'s) [62]. Here, we assume that impurities are distributed uniformly on the $z = 0$ plane with the sheet density of N_i. For a uniform distribution the average over all possible ensemble of impurities vanishes, so we approximate $\Phi(q,y)$ with its rms. After taking an average over y, including the screening, and in the limit of $W \ll W_s$, the interaction strength $[\mathcal{M}_{\text{ii}}(q) = \overline{\delta V}(q)]$ is

$$\mathcal{M}_{\text{ii}}(q) = \sqrt{\frac{N_i}{L}}\frac{-e^2\left[\int_{-\frac{W_s}{2}}^{\frac{W_s}{2}}dyK_0(|q\sqrt{y^2+d^2}|)^2\right]^{\frac{1}{2}}}{2\pi\varepsilon_0\kappa_b\varepsilon(q,\omega=0)}. \tag{C5}$$

Accordingly, the scattering weight is

$$\mathcal{W}^\pm_{k'-k,\text{ii}} = \frac{1}{4}\frac{N_ie^4}{4\pi^2\hbar\varepsilon_0^2\kappa_b^2}\frac{\int_{-\frac{W_s}{2}}^{\frac{W_s}{2}}dyK_0\left(|k-k'|\sqrt{y^2+d^2}\right)^2}{|\varepsilon(k-k',\omega=0)|^2}. \tag{C6}$$

3. Electron scattering with surface optical phonons

Applying the electromagnetic boundary conditions yields the dispersion relation of $\kappa_s(\omega)+1 = 0$, which its solutions are the surface optical modes, ω_{SO}. SO phonon modes for hBN and SiO_2 are provided in Ref. [29]. Assuming that the substrate has only one transverse optical mode, $\kappa_s(\omega) = \kappa_\infty + \frac{\omega^2_{\text{TO}}(\kappa_0-\kappa_\infty)}{\omega^2_{\text{TO}}-\omega^2}$, ω_{SO} and ω_{TO} are related by

$$\omega^2_{\text{SO}}(\kappa_\infty+1) = \omega^2_{\text{TO}}(\kappa_0+1). \tag{C7}$$

Now, we calculate the electric potential due to SO phonons. In the nonretarded regime, the electric potential can be written as

$$\Phi(\boldsymbol{r}) = \sum_{\boldsymbol{Q}} a_Qe^{-Q|z|+i\boldsymbol{Q}\cdot\boldsymbol{\rho}}, \tag{C8}$$

where $\boldsymbol{r} = \boldsymbol{\rho}+z\hat{z}$ (the cylindrical coordinates). $\boldsymbol{Q} = Q\hat{Q}$ is a two-dimensional vector on the xy plane, i.e., $\boldsymbol{Q} \equiv Q_x\hat{x}+Q_y\hat{y}$. The polarization, the dynamic equation of motion, and the electric displacement field in the crystal lattice are given

by

$$\boldsymbol{P}(\boldsymbol{r}) = e^* N_{\mathrm{uc},s} V_s^{-1} \boldsymbol{u}(\boldsymbol{r}), \tag{C9a}$$

$$-\mu\omega^2 \boldsymbol{u}(\boldsymbol{r}) = -\mu\omega_{\mathrm{TO}}^2 \boldsymbol{u}(\boldsymbol{r}) + e^* \boldsymbol{E}(\boldsymbol{r}), \tag{C9b}$$

$$\boldsymbol{D}(\boldsymbol{r}) \equiv \varepsilon_0 \kappa_s(\omega) \boldsymbol{E}(\boldsymbol{r}) = \varepsilon_0 \kappa_\infty \boldsymbol{E}(\boldsymbol{r}) + \boldsymbol{P}(\boldsymbol{r}). \tag{C9c}$$

V_s and $N_{\mathrm{uc},s}$ are the volume and the number of unit cells of the substrate, respectively. $\boldsymbol{u}(\boldsymbol{r})$ is the relative displacement of two adjacent atoms, with the reduced mass of μ. Incorporating Eqs. (C9c) and (C8) into Eq. (C9a), we obtain

$$\sum_{\boldsymbol{Q}} a_{\boldsymbol{Q}} Q(\hat{z} + i\hat{Q}) e^{Qz + i\boldsymbol{Q}\cdot\boldsymbol{\rho}} \Theta(-z) = \frac{e^* N_{\mathrm{uc},s} \boldsymbol{u}(\boldsymbol{r})}{V_s \varepsilon_0 (\kappa_\infty + 1)}. \tag{C10}$$

Considering the above equation, here is an ansatz for the displacement operator of mode $\boldsymbol{Q}$, $\mathbb{u}_{\boldsymbol{Q}}(\boldsymbol{r})$:

$$\mathbb{u}_{\boldsymbol{Q}}(\boldsymbol{r}) = \frac{u_{\boldsymbol{Q}} e^{Qz + i\boldsymbol{Q}\cdot\boldsymbol{\rho}}}{\sqrt{N_{\mathrm{uc},s}}} (\hat{z} + i\hat{Q}) \Theta(-z) \mathbb{B}_{\boldsymbol{Q}}, \tag{C11}$$

where $\Theta(z)$ is the heavyside step function, and $\mathbb{B}_{\boldsymbol{Q}} = b_{\boldsymbol{Q}} + b^\dagger_{-\boldsymbol{Q}}$. The total displacement is $\mathbb{u}(\boldsymbol{r}) = \sum_{\boldsymbol{Q}} \mathbb{u}_{\boldsymbol{Q}}(\boldsymbol{r})$. Yet, $u_{\boldsymbol{Q}}$ needs to be determined. The energy of each mode is $(2N_{\mathrm{SO}} + 1)\frac{\hbar\omega_{\mathrm{SO}}}{2}$. Also, the average energy of each mode over time is half the potential energy and half the kinetic energy. So for each mode of $\boldsymbol{u}_{\boldsymbol{Q}}(\boldsymbol{r})$ we can write [63–71]

$$\int d^3r \frac{1}{2} \varrho_s \omega_{\mathrm{SO}}^2 \langle \mathbb{u}_{\boldsymbol{Q}}(\boldsymbol{r}) \mathbb{u}_{\boldsymbol{Q}}{}^\dagger(\boldsymbol{r}) \rangle = \frac{1}{2}(2N_{\boldsymbol{Q}} + 1)\frac{\hbar\omega_{\boldsymbol{Q}}}{2}, \tag{C12}$$

ϱ_s is $\frac{N_{\mathrm{uc},s}\mu}{V_s}$. We solve the above equation for $u_{\boldsymbol{Q}}$, replace it in Eq. (C11), and get

$$\mathbb{u}(\boldsymbol{r}) = \sum_{\boldsymbol{Q}} \frac{\sqrt{\hbar Q} e^{Qz + i\boldsymbol{Q}\cdot\boldsymbol{\rho}}}{\sqrt{2\varrho_s A_s \omega_{\mathrm{SO}}}} (\hat{z} + i\hat{Q}) \Theta(-z) \mathbb{B}_{\boldsymbol{Q}}. \tag{C13}$$

Also, by using Eq. (C7), we solve Eqs. (C9) for e^* at $\omega = \omega_{\mathrm{SO}}$:

$$e^{*2} = \frac{\mu V_s \omega_{\mathrm{SO}}^2}{N_{\mathrm{uc},s}} \varepsilon_0 (\kappa_\infty + 1)^2 \left(\frac{1}{\kappa_\infty + 1} - \frac{1}{\kappa_0 + 1} \right). \tag{C14}$$

Incorporating Eq. (C14) and Eq. (C13) into Eq. (C10), we solve for a_q and substitute it in Eq. (C8) to get

$$\Phi(\boldsymbol{r}) = \sum_{\boldsymbol{Q}} \sqrt{\frac{2\pi\hbar}{e^2 A_s}} \frac{\Xi}{\sqrt{Q}} e^{-Q|z| + i\boldsymbol{Q}\cdot\boldsymbol{\rho}}, \tag{C15}$$

where $\Xi^2 = \frac{e^2\omega_{\mathrm{SO}}}{4\pi\varepsilon_0}(\frac{1}{\kappa_\infty+1} - \frac{1}{\kappa_0+1})$. Consequently, the interaction Hamiltonian is

$$\delta\mathbb{V}(\boldsymbol{r}) = \sum_{\boldsymbol{Q}} \sqrt{\frac{2\pi\hbar}{A_s}} \frac{\Xi}{\sqrt{Q}} e^{-Q|z| + i\boldsymbol{Q}\cdot\boldsymbol{\rho}} \mathbb{B}_{\boldsymbol{Q}}. \tag{C16}$$

We take the Fourier transform of $\delta\mathbb{V}(\boldsymbol{r})$ with respect to x at $z = d$, take an average over y, and change variables to obtain

$$\overline{\delta\mathbb{V}}(q) = \Xi \sqrt{\frac{2\pi\hbar}{A_s}} \sum_{q'} \frac{e^{-d\sqrt{q^2+q'^2}}}{\sqrt[4]{q^2+q'^2}} \mathrm{sinc}\left(\frac{q'W}{2\pi}\right) \mathbb{B}_{(q,q')}, \tag{C17}$$

where $\mathrm{sinc}(\frac{q'W}{2\pi}) = \sin(\frac{q'W}{2})(\frac{q'W}{2})^{-1}$. The equation of motion of any electronic operator is at least parabolic in terms of $\overline{\delta\mathbb{V}}(q)$. Also, the SO-phonon modes are assumed to be dispersionless, i.e., the number of SO phonons are independent of their wave vector. So, we can rewrite the interaction strength as

$$\mathcal{M}_{\mathrm{SO}}(q) = \sqrt{\frac{\hbar}{L}} \frac{\Xi}{\varepsilon(q,\omega=0)} \times \left[\int dQ_y \frac{e^{-2d\sqrt{q^2+Q_y^2}}}{\sqrt{q^2+Q_y^2}} \mathrm{sinc}^2\left(\frac{Q_y W}{2\pi}\right) \right]^{\frac{1}{2}}, \tag{C18}$$

and yet, nothing changes in the equation of motion of electronic operators. The $\mathrm{sinc}^2(\frac{q'W}{2\pi})$ term in Eq. (C18) is a window function which mimics the momentum conservation along the width of the ribbon. Unlike narrow ribbons, for wide ribbons (larger W) this window function narrows down and asymptotically goes to the Dirac delta function as $W \to \infty$. The scattering weight is

$$\mathcal{W}^{\pm}_{k'-k,\mathrm{SO}} = \frac{\Xi^2\left(N_{\mathrm{SO}} + \frac{1}{2} \pm \frac{1}{2}\right)}{2|\varepsilon(k-k',\omega=0)|^2} \times \left[\int dQ_y \frac{e^{-2d\sqrt{q^2+Q_y^2}}}{\sqrt{q^2+Q_y^2}} \mathrm{sinc}^2\left(\frac{Q_y W}{2\pi}\right) \right]. \tag{C19}$$

4. Electron scattering from line-edge roughness

The charge density can be assumed as $n = -en_0\Pi(\frac{y}{W})\delta(z)$, where n_0 is the surface carrier density, $\delta(\cdot)$ denotes the Dirac delta function, and $\Pi(.)$ denotes the rectangular function [1 for its argument being within (0,1), 1/2 for argument equal to 0 or 1, and zero elsewhere]. The change in the carrier density due to the line-edge roughness, $\Delta(x)$, at one edge is [72]

$$\delta n(\boldsymbol{r}) = \Delta(x)\frac{\partial n}{\partial y} = -en_0\Delta(x)\delta\left(y - \frac{W}{2}\right)\delta(z), \tag{C20}$$

which causes the electric potential energy of

$$\delta V(\boldsymbol{r}) = \iiint d^3\boldsymbol{r}' \frac{e^2 n_0 \Delta(x')\delta\left(y' - \frac{W}{2}\right)\delta(z')}{4\pi\varepsilon_0\kappa_b|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{C21}$$

We take the Fourier transform of $\delta V(\boldsymbol{r})$ with respect to x at $z = 0$, and then we take an average over y. So the interaction strength for the LER is

$$\mathcal{M}_{\mathrm{LER}}(q) = \frac{e^2 n_0}{\pi\varepsilon_0\kappa_b} \frac{\Delta(q)\int_{-\frac{W}{2}}^{\frac{W}{2}} dy K_0\left(q\left|y - \frac{W}{2}\right|\right)}{W\varepsilon(q,\omega=0)}, \tag{C22}$$

where $\Delta(q)$ is the Fourier transform of the LER function. Here, we consider an exponential correlation function:

$$\langle \Delta(x)\Delta(0) \rangle = \Delta^2 e^{-\frac{|x|}{\xi}}, |\Delta(q)|^2 = \frac{1}{L}\frac{2\Delta^2\xi}{1 + q^2\xi^2}, \tag{C23}$$

where Δ and ξ are the rms roughness and the correlation length, respectively. The scattering weight is

$$\mathcal{W}^{\pm}_{k'-k,\mathrm{LER}} = \frac{1}{4}\frac{e^4 n_0^2}{\hbar\pi^2\varepsilon_0^2\kappa_b^2}\left[\frac{2\Delta^2\xi}{1+(k-k')^2\xi^2}\right]\frac{\left[\int_0^1 d\eta K_0(|k-k'|W\eta)\right]^2}{|\varepsilon(k-k',\omega=0)|^2}. \tag{C24}$$

[1] S. A. Maier, M. L. Brongersma, P. G. Kik, S. Meltzer, A. A. Requicha, and H. A. Atwater, Adv. Mater. **13**, 1501 (2001).
[2] S. A. Maier and H. A. Atwater, J. Appl. Phys. **98**, 011101 (2005).
[3] D. K. Gramotnev and S. I. Bozhevolnyi, Nat. Photonics **4**, 83 (2010).
[4] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, Nat. Mater. **9**, 193 (2010).
[5] L. Novotny and N. Van Hulst, Nat. Photonics **5**, 83 (2011).
[6] V. M. Shalaev, Nat. Photonics **1**, 41 (2007).
[7] B. Luk'yanchuk, N. I. Zheludev, S. A. Maier, N. J. Halas, P. Nordlander, H. Giessen, and C. T. Chong, Nat. Mater. **9**, 707 (2010).
[8] S. Kawata, Y. Inouye, and P. Verma, Nat. Photonics **3**, 388 (2009).
[9] A. Alù and N. Engheta, Phys. Rev. E **72**, 016623 (2005).
[10] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, Nat. Mater. **7**, 442 (2008).
[11] A. Kabashin, P. Evans, S. Pastkovsky, W. Hendren, G. Wurtz, R. Atkinson, R. Pollard, V. Podolskiy, and A. Zayats, Nat. Mater. **8**, 867 (2009).
[12] H. A. Atwater and A. Polman, Nat. Mater. **9**, 205 (2010).
[13] N. Yu and F. Capasso, Nat. Mater. **13**, 139 (2014).
[14] Y. Fan, Z. Liu, F. Zhang, Q. Zhao, Z. Wei, Q. Fu, J. Li, C. Gu, and H. Li, Sci. Rep. **5**, 13596 (2015).
[15] Y. Fan, N.-H. Shen, T. Koschny, and C. M. Soukoulis, ACS Photonics **2**, 151 (2015).
[16] S. A. Maier, *Plasmonics: Fundamentals and Applications* (Springer Science & Business Media, New York, 2007).
[17] T. Low, A. Chaves, J. D. Caldwell, A. Kumar, N. X. Fang, P. Avouris, T. F. Heinz, F. Guinea, L. Martin-Moreno, and F. Koppens, Nat. Mater. **16**, 182 (2017).
[18] H. Yan, T. Low, W. Zhu, Y. Wu, M. Freitag, X. Li, F. Guinea, P. Avouris, and F. Xia, Nat. Photonics **7**, 394 (2013).
[19] A. Grigorenko, M. Polini, and K. Novoselov, Nat. Photonics **6**, 749 (2012).
[20] F. H. Koppens, D. E. Chang, and F. J. Garcia de Abajo, Nano Lett. **11**, 3370 (2011).
[21] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).
[22] M. Jablan, H. Buljan, and M. Soljačić, Phys. Rev. B **80**, 245435 (2009).
[23] X. Lin, N. Rivera, J. J. López, I. Kaminer, H. Chen, and M. Soljačić, New J. Phys. **18**, 105007 (2016).
[24] X. Lin, Y. Yang, N. Rivera, J. J. López, Y. Shen, I. Kaminer, H. Chen, B. Zhang, J. D. Joannopoulos, and M. Soljačić, Proc. Natl. Acad. Sci. **114**, 6717 (2017).
[25] X. Lin, I. Kaminer, X. Shi, F. Gao, Z. Yang, Z. Gao, H. Buljan, J. D. Joannopoulos, M. Soljačić, H. Chen *et al.*, Science Advances **3**, e1601192 (2017).
[26] F. J. Garcia de Abajo, ACS Photonics **1**, 135 (2014).
[27] A. C. Neto, F. Guinea, N. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[28] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen *et al.*, Nat. Nanotechnol. **6**, 630 (2011).
[29] F. Karimi, A. H. Davoody, and I. Knezevic, Phys. Rev. B **93**, 205421 (2016).
[30] J. Hu, W. Lu, and J. Wang, EPL (Europhysics Letters) **106**, 48002 (2014).
[31] P. B. Bennett, Z. Pedramrazi, A. Madani, Y.-C. Chen, D. G. de Oteyza, C. Chen, F. R. Fischer, M. F. Crommie, and J. Bokor, Appl. Phys. Lett. **103**, 253114 (2013).
[32] X. Wang, Y. Ouyang, X. Li, H. Wang, J. Guo, and H. Dai, Phys. Rev. Lett. **100**, 206803 (2008).
[33] T. Low and P. Avouris, ACS Nano **8**, 1086 (2014).
[34] V. W. Brar, M. S. Jang, M. Sherrott, S. Kim, J. J. Lopez, L. B. Kim, M. Choi, and H. Atwater, Nano Lett. **14**, 3876 (2014).
[35] A. Celis, M. Nair, A. Taleb-Ibrahimi, E. Conrad, C. Berger, W. de Heer, and A. Tejeda, J. Phys. D **49**, 143001 (2016).
[36] J. Bai, X. Duan, and Y. Huang, Nano Lett. **9**, 2083 (2009).
[37] Z. J. Qi, J. A. Rodríguez-Manzo, A. R. Botello-Méndez, S. J. Hong, E. A. Stach, Y. W. Park, J.-C. Charlier, M. Drndić, and A. C. Johnson, Nano Lett. **14**, 4238 (2014).
[38] L. Tapasztó, G. Dobrik, P. Lambin, and L. P. Biró, Nat. Nanotechnol. **3**, 397 (2008).
[39] G. Z. Magda, X. Jin, I. Hagymási, P. Vancsó, Z. Osváth, P. Nemes-Incze, C. Hwang, L. P. Biro, and L. Tapaszto, Nature (London) **514**, 608 (2014).
[40] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, Science **319**, 1229 (2008).
[41] J. Cai, P. Ruffieux, R. Jaafar, M. Bieri, T. Braun, S. Blankenburg, M. Muoth, A. P. Seitsonen, M. Saleh, X. Feng *et al.*, Nature (London) **466**, 470 (2010).
[42] A. Kimouche, M. M. Ervasti, R. Drost, S. Halonen, A. Harju, P. M. Joensuu, J. Sainio, and P. Liljeroth, Nat. Commun. **6**, 10177 (2015).
[43] R. Denk, M. Hohage, P. Zeppenfeld, J. Cai, C. A. Pignedoli, H. Söde, R. Fasel, X. Feng, K. Müllen, S. Wang *et al.*, Nat. Commun. **5**, 4253 (2014).
[44] G. D. Nguyen, F. M. Toma, T. Cao, Z. Pedramrazi, C. Chen, D. J. Rizzo, T. Joshi, C. Bronner, Y.-C. Chen, M. Favaro *et al.*, J. Phys. Chem. C **120**, 2684 (2016).
[45] P. Ruffieux, S. Wang, B. Yang, C. Sánchez-Sánchez, J. Liu, T. Dienel, L. Talirz, P. Shinde, C. A. Pignedoli, D. Passerone *et al.*, Nature (London) **531**, 489 (2016).
[46] L. Brey and H. A. Fertig, Phys. Rev. B **75**, 125434 (2007).
[47] M. Bagheri and M. Bahrami, J. Appl. Phys. **115**, 174301 (2014).
[48] C. E. P. Villegas, M. R. S. Tavares, G.-Q. Hai, and P. Vasilopoulos, Phys. Rev. B **88**, 165426 (2013).
[49] A. A. Shylau, S. M. Badalyan, F. M. Peeters, and A. P. Jauho, Phys. Rev. B **91**, 205444 (2015).
[50] D. R. Andersen and H. Raza, J. Phys.: Condens. Matter **25**, 045303 (2013).
[51] D. R. Andersen and H. Raza, Phys. Rev. B **85**, 075425 (2012).
[52] C. Vacacela Gomez, M. Pisarra, M. Gravina, J. M. Pitarke, and A. Sindona, Phys. Rev. Lett. **117**, 116801 (2016).
[53] A. Iurov, G. Gumbs, D. Huang, and V. M. Silkin, Phys. Rev. B **93**, 035404 (2016).

[54] T. Christensen, W. Wang, A.-P. Jauho, M. Wubs, and N. A. Mortensen, Phys. Rev. B **90**, 241414 (2014).
[55] I. Silveiro, J. M. P. Ortega, and F. J. G. De Abajo, Light: Sci. Appl. **4**, e241 (2015).
[56] S. Thongrattanasiri, A. Manjavacas, and F. J. García de Abajo, ACS Nano **6**, 1766 (2012).
[57] N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).
[58] Y.-W. Son, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **97**, 216803 (2006).
[59] D. Rodrigo, O. Limaj, D. Janner, D. Etezadi, F. J. G. de Abajo, V. Pruneri, and H. Altug, Science **349**, 165 (2015).
[60] R. Kundu, Mod. Phys. Lett. B **25**, 163 (2011).
[61] Y. Hancock, A. Uppstu, K. Saloriutta, A. Harju, and M. J. Puska, Phys. Rev. B **81**, 245402 (2010).
[62] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).
[63] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, J. Appl. Phys. **90**, 4587 (2001).
[64] J. I. Gersten, Surf. Sci. **92**, 579 (1980).
[65] J. J. Licari and R. Evrard, Phys. Rev. B **15**, 2254 (1977).
[66] N. Mori and T. Ando, Phys. Rev. B **40**, 6175 (1989).
[67] Z.-Y. Ong and M. V. Fischetti, Phys. Rev. B **86**, 165422 (2012).
[68] M. A. Stroscio and M. Dutta, *Phonons in Nanostructures* (Cambridge University Press, Cambridge, 2001).
[69] S. Q. Wang and G. D. Mahan, Phys. Rev. B **6**, 4517 (1972).
[70] L. Wendler, Phys. Status Solidi (b) **129**, 513 (1985).
[71] A. Betti, G. Fiori, and G. Iannaccone, Appl. Phys. Lett. **98**, 212111 (2011).
[72] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).