Plasmon-enhanced optical nonlinearity in graphene nanomeshes

F. Karimi,[*] S. Mitra,[†] S. Soleimanikahnoj,[‡] and I. Knezevic[§]
Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

(Received 3 June 2022; revised 7 June 2023; accepted 14 June 2023; published 20 July 2023)

Using the density-matrix formalism, we show that graphene nanomeshes (GNMs)—graphene sheets patterned with antidots—have large plasmon-enhanced nonlinear optical response. GNMs can be designed to behave as quasi-one-dimensional plasmonic crystals in which plasmons with large propagation lengths are efficiently excited. The associated third-order Kerr and third-harmonic-generation susceptibility can be as high as 10^{-7} and 10^{-9} $m^2\,V^{-2}$, respectively, over the mid-to-near-infrared frequency range. Furthermore, carrier-density tuning in GNMs can flip the propagation direction of plasmonic waves and enables bidirectional switching of optical signals.

DOI: 10.1103/PhysRevB.108.035414

I. INTRODUCTION

Nonlinear optics offers a promising platform for controlling and manipulating light at the nanoscale for nanophotonic applications [1–5]. Nonlinear optics relies on matter-mediated photon-photon interactions, which are intrinsically very weak. Hence, realizing nonlinear nanophotonics requires new materials and structures that will enhance nonlinear optical effects.

Graphene, the two-dimensional allotrope of carbon, offers a promising plasmonic platform for optoelectronic and photonic applications [6–8]. Being a semimetal, graphene has a lower electronic density of states and a lower number of free carriers per atom than metals [9]. As a result, plasmons in graphene are found in a lower frequency range (mid-to-near-infrared) than in metals (near-infrared to ultraviolet) [10–13]. This makes graphene a more suitable platform for applications in the telecom spectrum [6–8]. Another advantage of graphene with respect to metals is that the graphene carrier density and, consequently, its optoelectronic properties can be electrically tuned by an external gate voltage. Graphene supports electrically tunable sheet carrier densities of up to $n_s = 10^{14}$ cm^{-2}, or equivalently, the Fermi energies of up to $E_F = \hbar v_F \sqrt{n_s} \approx 1.2$ eV, with the Fermi velocity $v_F = 10^6$ $m\,s^{-1}$ [14–16].

Graphene and its nanostructures strongly interact with light and are a fitting class of materials for nonlinear optics applications [17,18]. It is therefore not surprising that graphene nonlinear optics and graphene plasmonics have both been very active areas of research in recent years [6,19,20]. A proven way to enhance light-matter interaction and nonlinear response is by integration of graphene with systems that support photonic resonances [1,21–27]. An alternative is through plasmonic field enhancement combined with quantum confinement [28], which is attractive because of electrical tunability of the carrier density (and thus plasmonic resonances) by a back gate. For example, plasmon-enhanced Kerr nonlinearity of small graphene nanoislands with sub-10-nm features could be as large as 10^{-11} $m^2\,V^{-2}$, almost four orders of magnitude greater than in graphene sheets [18,29]. However, the synthesis of sub-10-nm nanoislands with precise geometry, size, and edge termination, which is required for the control of their plasmonic response [18], is very challenging and makes these systems difficult to envision in integrated nanophotonics applications.

In contrast, the synthesis of graphene sheets and integration of them into nanophotonics and nanoelectronics are mature, well-developed processes [30,31]. However, launching propagating plasmonic waves in graphene is challenging in practice because the wave vectors of graphene plasmons are much larger than the wave vector of the free-space electromagnetic waves with the same frequency. Owing to the wave vector mismatch, graphene plasmons cannot be simply excited by free-space light. Fortunately, one can launch propagating plasmonic waves in graphene sheets and ribbons patterned with periodically distributed antidots (holes). The periodicity aids with momentum conservation, which enables plasmon excitation. Successful plasmon excitation in these structures has been experimentally demonstrated [32–36].

In this paper, we show that graphene nanomeshes (GNMs)—graphene sheets patterned with antidots—offer a promising core material for nonlinear optics applications. Figure 1 shows the schematic of our proposed graphene nanomesh sitting on the hexagonal boron nitride (hBN) substrate. The GNM can be fabricated by etching a rectangular antidot superlattice out of the supported graphene sheet. The antidot diameter is d and superlattice periods are a and $10a$ along the x and y axes, respectively. We chose a periodicity of $a = 100$ nm and the antidot diameter of $d = 30$ nm. Fabrication of such antidot superlattices is entirely feasible using existing experimental techniques [30,37]. We show that properly designed GNMs support plasmonic waves with long propagating lengths in the mid-to-near-infrared frequency range. Owing to the plasmonic field enhancement, GNMs have a dynamically tunable, broadband, and strong nonlinear plasmonic response. The GNM third-harmonic-generation

[*]karimi2@wisc.edu
[†]smitra8@wisc.edu
[‡]soleimanikah@wisc.edu
[§]iknezevic@wisc.edu

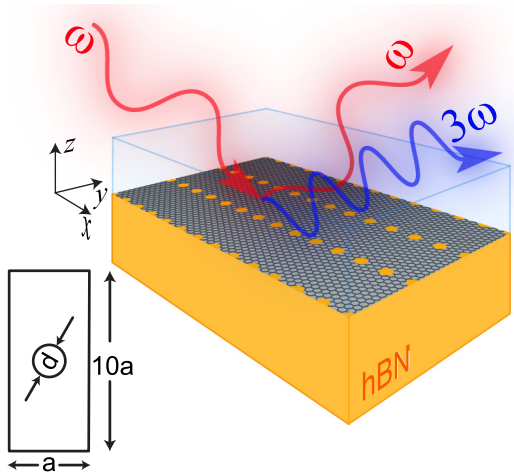

FIG. 1. Schematic of a graphene nanomesh, placed on an hBN substrate. Incident light with in-plane TM polarization (meaning nonzero electric-field component along the plasmon-propagation direction) causes linear as well as third-order nonlinear optical responses. Inset: The unit cell of the rectangular superlattice: The lattice constants along the x and y axes are a and $10a$, respectively. The antidot diameter is d.

and third-order Kerr susceptibilities can be as high as 10^{-9} and 10^{-7} $\mathrm{m^2\,V^{-2}}$, respectively. The latter values far exceed the Kerr susceptibilities of 10^{-13} to 10^{-11} $\mathrm{m^2\,V^{-2}}$ reported in other two-dimensional systems such as graphdiyne [38], black phosphorus [39], tin sufide [40], and tellurium-based devices [41]. Our findings reveal the capability of quasi-one-dimensional GNMs for nonlinear nanophotonic applications, particularly bidirectional switching and modulation of optical signals.

The rest of the paper is organized as follows. In Sec. II, we overview the theoretical model and discuss the dominant scattering mechanisms in the different carrier-density ranges. In Sec. III, we focus on the plasmonic and optical properties of the GNMs. We conclude in Sec. IV.

II. THEORETICAL MODEL

The plasmonic excitations of the GNM in Fig. 1(a) are calculated using the density-matrix approach [17,42,43] in response to a TM-polarized incident light guided along the x direction. The time evolution of the electronic density matrix $\rho_e(t)$ in the Schrödinger picture reads

$$\frac{d\rho_e(t)}{dt} = -\frac{i}{\hbar}[\mathbb{H}_e + \mathbb{V}_{\mathrm{SCF}}(t), \rho_e(t)] + \frac{[\rho_e(t) - \rho_0]}{\tau_F}. \quad (1)$$

Here, ρ_0 is the equilibrium density matrix, $\mathbb{H}_e$ is the unperturbed electronic Hamiltonian, $\mathbb{V}_{\mathrm{SCF}}(t)$ is the self-consistent field, and τ_F denotes the electron relaxation time. To accurately calculate τ_F, we account for electron scattering via intrinsic phonons, ionized impurities, surface optical (SO) phonons of the hBN substrate, and antidot-edge roughness (AER). For the GNMs of interest, the antidot area is less than 1% of the unit cell area. Given the relatively tiny area of antidots, we could accurately approximate the GNM band structure with the graphene band structure. Hence, the energy dispersion is approximated as linear and isotropic $E_{\boldsymbol{k}l} = \hbar v_F|\boldsymbol{k}|$, with $l = 1$ and $l = -1$ for the conduction and valence bands, respectively. We perturbatively solve Eq. (1) for the density matrix via a similar procedure as in Ref. [17], and calculate the surface polarization $P^{(s,p_s)}$, with s and p_s denoting the response order and the corresponding harmonic, respectively. Next, we calculate the macroscopic quantities such as the linear dielectric function ε, loss function $\sigma_{\mathrm{abs}} = -\mathrm{Im}\{1/\varepsilon\}$ (which measures field enhancement), and the sth-order p_sth-harmonic nonlinear susceptibility with respect to the external field ($\chi^{(s,p_s)}$). We obtain the plasmon dispersion by seeking the peaks of the loss function. The plasmon propagation length (L_p) equals $1/\Delta q$, where Δq is the half width at half maximum of the loss-function peak in the wave vector direction [13]. The numerical procedure to solve Eq. (1), calculation of scattering rates, and optical properties are given in Appendices A and B. While the derivations of most relevant scattering rates were published previously in [10], the electron scattering rate from antidot edge roughness (AER) is new and had not been published before; details of this derivation can be found in Appendix B.

In order to have pronounced plasmonic effects, we need to efficiently couple the diffracted wave and plasmon modes and excite plasmon modes with sufficiently long propagation length. To start, it is necessary to have a nonzero component of the electric field along the plasmon wave vector (see Appendix A); if the sheet were considered a planar plasmonic waveguide, incident light would have to be TM polarized. While our calculation does not require a specific way of achieving such excitation, it could be achieved, for example, by free-space TEM light at oblique incidence. Furthermore, optimizing the plasmon propagation length and, consequently, the plasmonic field enhancement requires a thorough knowledge of the contributing plasmon damping pathways. A key damping pathway is Landau damping, which happens when a plasmon decays to generate an electron-hole pair. Increasing E_F pushes the onset of Landau damping toward higher frequencies [11,12]. To completely switch off the Landau damping up to the telecom frequency range, the Fermi energy should be $\sim$1 eV (or $n_s = 7 \times 10^{13}$ $\mathrm{cm^{-2}}$) [11]. As mentioned earlier, graphene's carrier density can be electrically tuned to these high values by a back-gate bias voltage.

In the absence of Landau damping, other important electron scattering mechanisms are intrinsic phonons, ionized impurities, AER, and SO phonons of the hBN substrate. Electron scattering with intrinsic phonons increases parabolically with electron energy (see Appendix B). Therefore, at room temperature and high Fermi energies of interest ($\sim$1 eV), screened intrinsic phonons become the dominant dissipative mechanism and significantly diminish the plasmon propagation lengths. However, screened intrinsic-phonon scattering decreases with decreasing temperature, and so does its detrimental effect on L_p. Figure 2 shows the electron scattering rates, $\tau^{-1}(E)$, for different scattering mechanisms for $E_F =$ 1 eV at 77 K. τ^{-1} is calculated as a function of energy with respect to the bottom of the conduction band. We chose very conservative values for the relevant scattering parameters. We assume the impurity density 4×10^{11} $\mathrm{cm^{-2}}$, an order of magnitude larger than in the typical graphene-on-hBN devices [44]. The AER is assumed to be exponentially

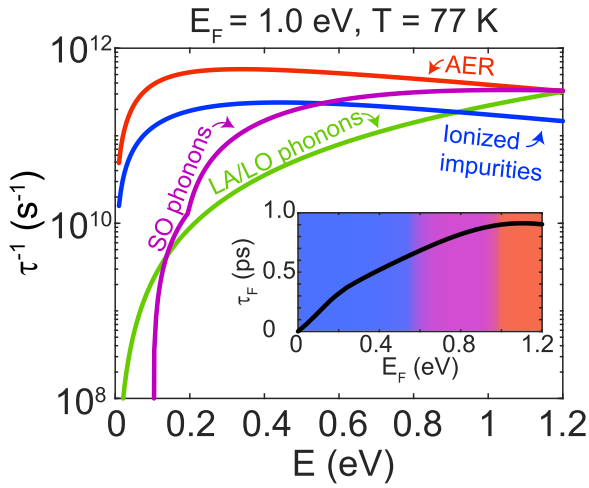

FIG. 2. Electron relaxation rates for different scattering mechanisms vs energy with respect to the bottom of the conduction band for a doped GNM with $E_F = 1.0$ eV at 77 K. For the GNM, $a = 100$ nm and $d = 30$ nm. The active scattering mechanisms are intrinsic phonons (LA/LO phonons), ionized impurities, antidot-edge roughness (AER), and surface optical (SO) phonons of the hBN substrate. The impurity density is 4×10^{11} cm^{-2}. The AER is assumed to be exponentially correlated with an rms roughness of 2 nm and a correlation length of 3 nm. Inset: The GNM electron relaxation time as a function of the Fermi energy. At low (blue-shaded), intermediate (magenta-shaded), and high (red-shaded) E_F, ionized impurities, SO phonons, and AER are the dominant scattering mechanisms, respectively. The color of a region in the inset corresponds to the color of the scattering-rate curve for the dominant mechanism in the main panel.

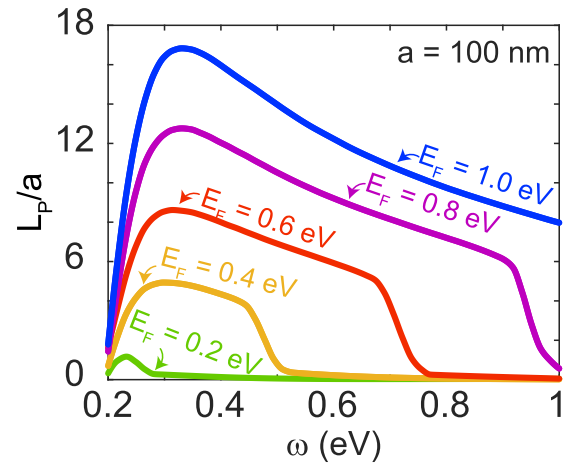

FIG. 3. Normalized plasmon propagation length for different Fermi energies. For lower Fermi energies, Landau damping occurs at lower frequencies and results in a massive decrease in the plasmon propagation lengths.

correlated, with an rms roughness of 2 nm and a correlation length of 3 nm. As can be seen, AER scattering is the dominant dissipative mechanism for $E_F = 1$ eV. However, at lower values of the Fermi energy, other scattering mechanisms dominate. The inset of Fig. 2 shows the electron relaxation time $\tau_F \equiv \tau(E_F)$ as a function of the Fermi energy. At low carrier densities, ionized impurities are the dominant scattering source. If we increase the carrier density up to intermediate values, SO phonons become the major scattering mechanism. At high carrier densities, as seen earlier, AER scattering is the major dissipative mechanism. The AER scattering rate increases with carrier density but changes negligibly with temperature. Given the conservative values used to calculate τ_F, the electron relaxation time is ∼0.8 ps for $E_F \approx 1$ eV.

III. PLASMONIC AND OPTICAL RESPONSE OF GRAPHENE NANOMESHES

Given the electron relaxation time, we can calculate the plasmon propagation length. Figure 3 shows the normalized plasmon propagation length L_p/a along the x direction as a function of frequency for different Fermi energies. Since the highest optical phonon mode of hBN is 195 meV, we are interested in the frequencies greater than about 200 meV in order to avoid plasmon suppression due to the coupling between plasmons and the substrate SO phonons [10]. With increasing frequency, Landau damping kicks in and hinders plasmon propagation, and requires higher carrier densities to counter. In Fig. 3, the dips in L_p/a mark the onset of Landau damping. For $E_F = 1$ eV, Landau damping is negligible in the frequency range of interest, and we have L_p/a between 8 to 17, with an average value of 11.8. These long plasmon propagation lengths imply efficient diffraction coupling and excitation of plasmonic waves. In contrast to what is happening in the x direction, plasmons barely travel two lattice constants ($20a$) in the y direction and decay too fast to benefit from the periodic pattern. Since plasmons experience the GNM in the y direction the same as an unpatterned graphene sheet, the diffraction coupling is extremely weak to launch plasmons in this direction. In other words, our designed GNM behaves as a quasi-one-dimensional (quasi-1D) plasmonic crystal. The 2D GNM with antidots behaves as quasi-1D plasmonic crystals by virtue of different periodicity along different axes. A key advantage of quasi-1D plasmonic crystals over 2D ones is that plasmonic waves are launched coherently in the predetermined direction.

Now, given the appropriately designed GNM, we can investigate its optical properties by calculating the loss function (σ_{abs}). Figure 4(a) shows the loss function as a function of E_F and frequency. The bright narrow strips correspond to plasmon resonances. A plasmon is excited when the difference between the wave vector of the illumination and the wave vector of the plasmon is an integer multiple of a reciprocal-lattice vector's magnitude (i.e., $2\pi n/a$, with n denoting the diffraction order). For each diffraction order, there is a plasmon branch comprising a pair of bright narrow strips: Forward-propagating (FWD) and backward-propagating (BKWD) modes along the x direction. [The splitting of forward and backward propagation comes from the asymmetry with respect to the polarization of the input light. There are terms involving a dot product between the plasmon wave vector and the in-plane electric field that are key for higher-order field induction; see Eqs. (A4) and (A6) in Appendix A 1. If plasmons are excited by free-space light, this would require oblique incidence.] The analytical

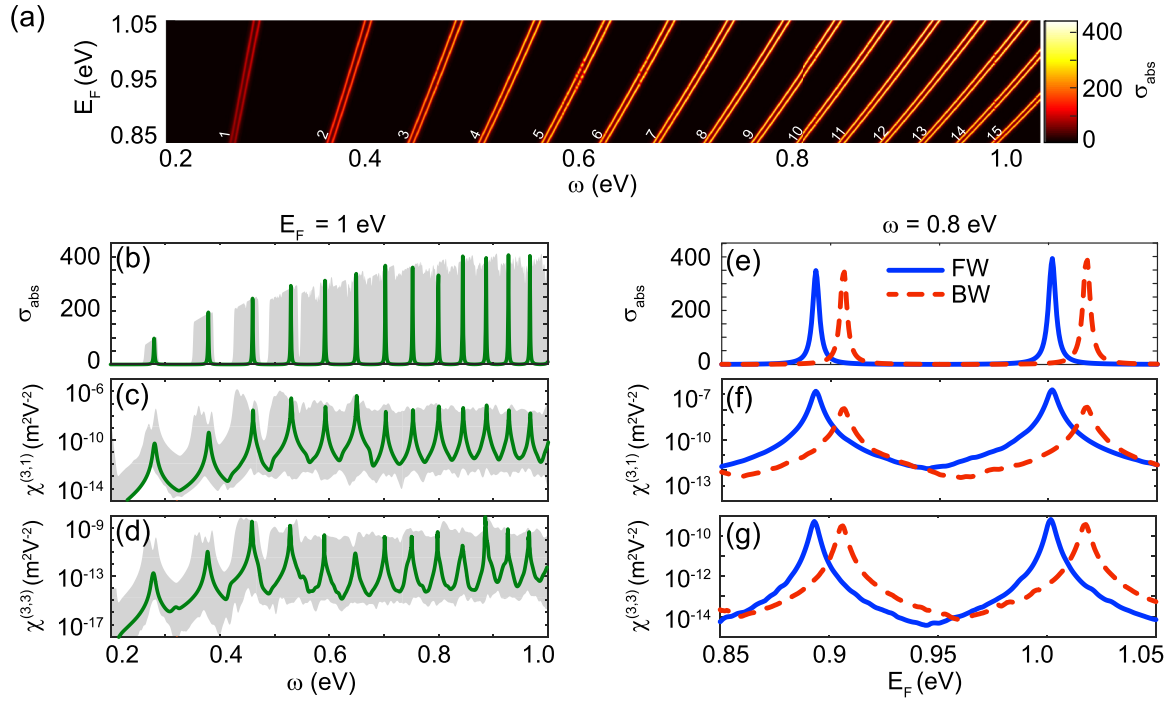

FIG. 4. (a) The GNM loss function as a function of the Fermi energy and frequency. Each branch comprises two modes: Forward- and backward-propagating waves. (b) Loss function, (c) magnitude of the third-order Kerr susceptibility, and (d) magnitude of the third-harmonic-generation (THG) susceptibility for a fixed Fermi energy ($E_F = 1$ eV) as a function of frequency for forward-propagating modes. The gray-shaded areas show the corresponding quantities calculated for the Fermi energy in the range of 0.85 to 1.05 eV. (e) Loss function, (f) magnitude of third-order Kerr susceptibility, and (g) magnitude of the THG susceptibility for a fixed frequency ($\omega = 0.8$ eV), as a function of the Fermi energy for forward-propagating (blue) and backward-propagating (red) modes.

plasmon dispersion is $n = \xi a \omega^2$, where $\xi = \hbar^2 \epsilon_0 \epsilon_b / (e^2 E_F)$ and ϵ_b denotes the background relative permittivity. (The background permittivity is defined as the average of the values for the top and bottom dielectrics. In our case, as the top half-domain is air, $\epsilon_b = \frac{1+\epsilon_{hBN}}{2}$ [10].) Based on the analytical plasmon dispersion, the frequency separation of two consecutive branches (e.g., n and $n+1$) equals $\Delta\omega = (\sqrt{n+1} - \sqrt{n})/\sqrt{\xi a}$, which decreases at higher diffraction orders as well as higher Fermi energies. Within a branch and for a fixed Fermi energy, the BKWD and FWD modes are separated by $\delta\omega$, with the BKWD mode occurring at a smaller frequency. In the limit of $\delta\omega \ll \omega$, $\delta\omega$ is independent of frequency and equals $(2\pi c \xi)^{-1}$. For the Fermi energies and frequencies of interest ($E_F \sim 1$ eV), $\delta\omega$ is ~5 meV.

After characterizing the plasmonic response of the GNM, we calculate its nonlinear optical response. Figures 4(b)–4(d) show the loss function, third-order Kerr susceptibility ($\chi^{(3,1)}$), and third-harmonic-generation (THG) susceptibility ($\chi^{(3,3)}$) for the FWD modes as a function of frequency and for a fixed E_F. For $E_F = 1$ eV, $\chi^{(3,1)}$ and $\chi^{(3,3)}$ are as high as 10^{-7} and 10^{-9} m^2 V^{-2}, respectively. The loss function in Fig. 4(b) quantifies plasmonic field enhancement, and can be as high as 300–400 at resonance. While plasmon enhancement increases the GNM optical nonlinearity to unprecedented large values, its effect is narrowband. Owing to the long propagation length of plasmons, the loss-function peaks at plasmon resonances are narrow and, therefore, plasmons significantly enhance the nonlinear optical response but over a narrow frequency range. The solution to broaden the nonlinear optical response is tuning the plasmon resonances by changing the carrier density by changing the back-gate voltage. The gray-shaded areas in Figs. 4(b)–4(d) represent the corresponding quantities calculated for Fermi energies in the range of 0.85 to 1.05 eV, or, equivalently, the carrier density range of 5×10^{13} to 8×10^{13} cm^{-2}. By tuning the carrier density over this small range, very large optical nonlinearity (as high as 10^{-7} m^2 V^{-2} for $\chi^{(3,1)}$ and 10^{-9} m^2 V^{-2} for $\chi^{(3,3)}$) is achieved over a broad frequency range. The tunable, broadband, and strong nonlinear plasmonic response makes GNMs an excellent platform for the modulation of optical signals.

It should be noted that the GNM-on-hBN device also provides high modulation speed. In the case of our graphene-based nanomesh, because the carrier mobility is exceptionally high, the modulation speed is not limited by the carrier transit time. In fact, the limiting factor is the parasitic response of the device. For instance, the experimentally measured values for the unity-power gain frequencies, $f_{\max}$, of graphene-hBN devices are ~10 GHz [45,46].

In addition to modulation applications, GNMs offer impressive capabilities for switching applications. Figures 4(e)–4(g) show the GNM σ_{abs}, $\chi^{(3,1)}$, and $\chi^{(3,3)}$ as a function of the Fermi energy at the fiber-optics frequency ($\omega = 0.8$ eV, or equivalently, 1.55 μm) and for both FWD and BKWD modes. (See the Supplemental Material for breakdown into real and imaginary parts [47].) For a given diffraction order and frequency, the FWD mode is excited at a lower Fermi energy. By fine-tuning the Fermi energy, which is readily achieved by a back gate, not only could we control the strength

of the GNM plasmon and nonlinear optical responses, but we could also determine the direction of the plasmonic wave. The latter capability enables bidirectional switching of optical signals by small changes in the Fermi level, basically moving the system between FWD and BKWD peaks. It should be emphasized that, given their frequency separation of ∼5 meV and carrier-density separation of $\sim 3 \times 10^{12}$ cm^{-2}, the FWD and BKWD modes can be completely resolved.

It is worth emphasizing that the above calculations were all done for graphene on hBN. For completeness, we also calculated third-order response of graphene on two other common substrates, the polar SiO_2 and the nonpolar diamondlike carbon (DLC) [10], and all third-order susceptibilities are comparable to one another. (See the Supplemental Material for the role of different substrates [47].) The main reason is that the AER scattering is the dominant scattering mechanism in all systems (it overshadows SO scattering in SiO_2, and DLC has no SO scattering). Therefore, the susceptibilities of the nanomesh do not change drastically on different substrates. However, SO phonon modes are critical in determining the operating-frequency range, given that SO phonon modes set the lower limit of the excitation frequency. Furthermore, the energy separation between the FWD and BKWD modes depends on the substrate (see above for a $\delta\omega$ estimate).

IV. CONCLUSION

In summary, we showed that GNMs have a large plasmon-enhanced nonlinear optical response. We designed a GNM that behaves as a quasi-one-dimensional plasmonic crystal in which plasmons with large propagation lengths are efficiently excited. The periodicity of the GNM can be optimized for efficient diffraction coupling at given frequencies. We showed that the GNM has tunable, broadband, and strong nonlinear plasmonic response. The third-order Kerr and THG susceptibility can be as high as 10^{-7} and 10^{-9} m^2 V^{-2}, respectively, over the mid-to-near-IR range. Moreover, by fine-tuning the carrier density, we could also switch the direction of the plasmonic wave. These findings suggest the GNM as a core material for integrated nanophotonic applications, particularly for switching and modulation of optical signals.

ACKNOWLEDGMENTS

The work was primarily supported by the U.S. Department of Energy Award No. DE-SC0023178 (TCMP program). Preliminary work was supported by the Splinter Professorship and the U.S. Department of Energy Award No. DE-SC0008712 (PBM program). This work was performed using the computing resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences.

APPENDIX A: CALCULATION OF THE GNM NONLINEAR PLASMONIC RESPONSE

1. The self-consistent-field approximation

Here, we use the perturbation theory to calculate the graphene nanomesh (GNM) third-order susceptibility. We assume an incident field as $V_{\rm inc}(t) = Ve^{i\boldsymbol{q}_1\cdot\boldsymbol{r}-i\omega t}$. $\boldsymbol{r}$ is the position vector in the $x-y$ plane. The GNM periodic pattern diffracts the incident light. If a diffracted mode couples with a plasmon mode, the self-consistent field reads $V_{\rm SCF}(t) = V_{\rm SCF}e^{i\tilde{\boldsymbol{q}}\cdot\boldsymbol{r}-i\omega t}$, where $\tilde{\boldsymbol{q}}$ is the wave vector of the excited plasmon corresponding to ω and $\boldsymbol{q}$. The induced carrier density can be written as

$$n(\boldsymbol{r}, z, t) = \delta(z)\sum_{s,p_s} n_{\rm ind}^{(s,p_s)} e^{p_s(i\tilde{\boldsymbol{q}}\cdot\boldsymbol{r}-i\omega t)}, \tag{A1}$$

where s is a natural number, denoting the perturbation order, and p_s is the harmonics order. Given the expansion of the induced carrier density, the inhomogeneous wave equation for the induced potential energy, $V_{\rm ind}^{(s,p_s)}(z)$, reads

$$\left[\frac{\partial}{\partial^2 z} + \left(iQ_{p_s}\right)^2\right] V_{\rm ind}^{(s)}(p_s\omega, z) = -\frac{-e}{\varepsilon_b\varepsilon_0} n^{(s,p_s)}\delta(z), \tag{A2}$$

where $(iQ_{p_s})^2 = \frac{\varepsilon_b(p_s\omega)^2}{c^2} - |p_s\tilde{\boldsymbol{q}}|^2$. In the nonretarded regime, we can assume $Q_{p_s} \approx |p_s\tilde{\boldsymbol{q}}|$. By solving the above equation for $V_{\rm ind}^{(s,p_s)}(z)$, we obtain

$$V_{\rm ind}^{(s,p_s)}(z) = \frac{e^2}{\varepsilon_b(p_s\omega)\varepsilon_0}\frac{e^{-Q_{p_s}|z|}}{2Q_{p_s}} n^{(s,p_s)}. \tag{A3}$$

Since we assumed that the quasi-two-dimensional (quasi-2D) system is lying in the $z = 0$ plane, we calculate all quantities at $z = 0$ and, to simplify the notation, from now on we drop the z argument. We also assume incident in-plane polarization to be along the x axis. By knowing that $-e\boldsymbol{E}_{\rm ind}^{(s,p_s)} = -i(p_s\tilde{\boldsymbol{q}})V_{\rm ind}^{(s,p_s)}$, we rewrite Eq. (A3) in terms of the electric fields:

$$E_{\rm ind}^{(s,p_s)} = \frac{-i(p_s\tilde{\boldsymbol{q}})\cdot\boldsymbol{e_x}}{\varepsilon_b(p_s\omega)\varepsilon_0}\frac{P^{(s,p_s)}}{2Q_{p_s}}\left(\frac{e}{\tilde{q}}\right)^s E^{\frac{s+p_s}{2}} E^{*\frac{s-p_s}{2}}, \tag{A4}$$

where we defined polarization as

$$P^{(s,p_s)} = \frac{-en^{(s,p_s)}}{V_{\rm SCF}^{\frac{s+p_s}{2}} V_{\rm SCF}^{*\frac{s-p_s}{2}}}. \tag{A5}$$

(Note the $\tilde{\boldsymbol{q}}\cdot\boldsymbol{e_x}$ term stemming from the incident field's in-plane polarization along the x direction.) From Eq. (A4), one can simply obtain an expression for the nonlinear optical susceptibility in response to the incident field:

$$\chi^{(s,p_s)} = \frac{-i(p_s\tilde{\boldsymbol{q}})\cdot\boldsymbol{e_x}}{\varepsilon_b(p_s\omega)\varepsilon_0}\frac{P^{(s_1+s_2)}(p_1\omega_1 + p_2\omega_2)}{2\tilde{\varepsilon}^{(s,p_s)}Q_{p_s}}\left(\frac{e}{\tilde{q}}\right)^s, \tag{A6}$$

where $\tilde{\varepsilon}^{(s,p_s)} \equiv \varepsilon(\tilde{q},\omega_1)^{\frac{s+p_s}{2}}\varepsilon^*(\tilde{q},\omega_1)^{\frac{s-p_s}{2}}$ and $\tilde{q}(\omega)$ is the diffracted wave vector that results in the maximum loss function among all the diffracted wave vectors. The surface polarization $P^{(s_1+s_2)}$, needed to obtain the linear as well as nonlinear susceptibility is calculated next, based on a master-equation formalism.

2. The master-equation formalism

We assume $\mathbb{H}_e$ to be the unperturbed Hamiltonian of free electrons in the GNM lattice, with its eigenkets and eigenenergies being represented by $|\boldsymbol{k}l\rangle$ and $\epsilon_{\boldsymbol{k}l}$, respectively. $\boldsymbol{k}$ is the in-plane electron wave vector and l is the band index. The sth-order and p_sth-harmonic induced charge density and

polarization in the second-quantization representation are

$$n_{\rm ind}^{(s,p_s)} = \frac{1}{A}\sum_{\boldsymbol{k},l',l}\left\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\right\rangle^{(s,p_s)}(\boldsymbol{kl}|\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'),$$
$$P^{(s,p_s)} = -\frac{e}{A}\sum_{\boldsymbol{k},l',l}\frac{\left\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\right\rangle^{(s,p_s)}(\boldsymbol{kl}|\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l')}{V_{\rm SCF}^{\frac{s+p_s}{2}}V_{\rm SCF}^{*\,\frac{s-p_s}{2}}}, \tag{A7}$$

where c and $c^{\dagger}$ are the electronic creation and destruction operators, respectively, and $(\boldsymbol{k}'l'|\boldsymbol{kl}) \equiv \langle \boldsymbol{k}'l'|\exp[-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}]|\boldsymbol{kl}\rangle$. Now, we derive a quantum-master equation and perturbatively solve it for the higher-order coherence terms, i.e., $\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\rangle^{(s,p_s)}$. The total Hamiltonian within the self-consistent-field approximation is

$$\mathbb{H}(t) = \mathbb{H}_{\rm e} + \mathbb{V}_{\rm SCF}(t), \tag{A8}$$

where $\mathbb{V}_{\rm SCF}(t) = \sum_{\boldsymbol{k},l',l} V_{\rm SCF}e^{-i\omega t}(\boldsymbol{k}+\tilde{\boldsymbol{q}}l'|\boldsymbol{kl})c_{\boldsymbol{k}+\tilde{\boldsymbol{q}}l'}^{\dagger}c_{\boldsymbol{kl}} + \text{H.c.}$ In the Schrödinger picture, the equation of motion for the density matrix is

$$\frac{d\rho_e(t)}{dt} = -\frac{i}{\hbar}[\mathbb{H}_{\rm e},\rho_e(t)] - \frac{i}{\hbar}[\mathbb{V}_{\rm SCF}(t),\rho_e(t)] - \frac{1}{\tau_F}[\rho_e(t)-\rho_e(0)], \tag{A9}$$

where τ_F is the electron relaxation time. Now, we use a perturbative approach to solve Eq. (A9) for the density operator. The perturbation expansion of $\rho_e(t)$ is $\sum_{sp_s}\rho_e^{(s,p_s)}e^{-i(p_s\omega)t} + \text{H.c.}$ By substituting the perturbation expansion of the density matrix into Eq. (A9), using the mean-field approximation, and seeking the harmonic solutions, the equation of motion for $\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\rangle^{(s,p_s)}$ would be

$$\begin{aligned}\hbar p_s\omega_1\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\rangle^{(s,p_s)} &= (\epsilon_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}-\epsilon_{\boldsymbol{kl}})\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\rangle^{(s,p_s)} - \frac{i\hbar}{\tau_F}\langle c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\rangle^{(s,p_s)}\\ &+ V_{\rm SCF}\sum_{\boldsymbol{k}'mm'}\text{tr}_e\left\{\left[c_{\boldsymbol{k}'+\tilde{\boldsymbol{q}}m'}^{\dagger}c_{\boldsymbol{k}'m},\rho_e^{(s-1,p_s-1)}\right]c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\right\}(\boldsymbol{k}'+\tilde{\boldsymbol{q}}m'|\boldsymbol{k}'m)\\ &+ V_{\rm SCF}^{*}\sum_{\boldsymbol{k}'mm'}\text{tr}_e\left\{\left[c_{\boldsymbol{k}'-\tilde{\boldsymbol{q}}m'}^{\dagger}c_{\boldsymbol{k}'m},\rho_e^{(s-1,p_s+1)}\right]c_{\boldsymbol{kl}}^{\dagger}c_{\boldsymbol{k}+p_s\tilde{\boldsymbol{q}}l'}\right\}(\boldsymbol{k}'-\tilde{\boldsymbol{q}}m'|\boldsymbol{k}'m).\end{aligned} \tag{A10}$$

By discretizing the Brillouin zone and employing the vectorial form, Eq. (A10) could be solved numerically. It should be noted that the quartic terms of creation/destruction operators can be simplified to multiplication of two quadratic terms via Wick's theorem and mean-field approximation [10].

APPENDIX B: SCATTERING RATES

To accurately calculate τ_F, we account for electron scattering via intrinsic phonons, ionized impurities, SO phonons of the hBN substrate, and antidot edge roughness (AER). The expression for the SO-phonon and ionized-impurity scattering rates are provided in Ref. [10]. The screened intrinsic phonon scattering rate reads [16]

$$\Gamma_m(E) = \frac{8}{\pi}\frac{D_{ph}^2}{\varrho v_s}\frac{E^2}{\hbar^3 v_F^3}\int_0^1\frac{\Theta_{BG}}{T}\frac{x^4\sqrt{1-x^2}}{(e^{x\Theta_{BG}/T}-1)^2}dx, \tag{B1}$$

where T is the lattice temperature and $\Theta_{BG} = 2E_F v_s/(v_F k_B)$ is the Bloch-Gruneisen temperature. $D_{ph} = 25$ eV is the deformation potential and $v_s = 2\times10^4$ m s^{-1} denotes the sound velocity in graphene.

Antidot edge roughness: Interaction Hamiltonian and scattering rate

Here, we derive the interaction Hamiltonian and scattering rate of electrons and antidot edge roughness (AER) in a quasi-2D electronic system. The charge density can be assumed as

$$\rho = -en_s\delta r(z), \tag{B2}$$

where n_s is the surface carrier density. The carrier density's variation due to the edge roughness of an antidot centered at the origin with radius of $r_0 = d/2$, in the cylindrical coordinates, is

$$\delta\rho(\boldsymbol{r},z) = \delta r(\theta)\frac{\partial n}{\partial r} = -en_s\delta r(\theta)\delta r(r-r_0)\delta r(z), \tag{B3}$$

where $\delta r(\theta')$ denotes the edge roughness of the antidot. The electric potential due to $\delta\rho(\boldsymbol{r})$ of the antidot is

$$\delta\Phi_{\boldsymbol{0}}(\boldsymbol{r}) = \frac{-en_s}{4\pi\varepsilon_s\kappa_b}\iiint d^2r'dz'\frac{1}{|\boldsymbol{r}-\boldsymbol{r'}|}\delta r(\theta')\delta r(r'-r_0)\delta r(z'). \tag{B4}$$

The $\boldsymbol{0}$ subscript denotes that the antidot is centered at the origin. Taking the Fourier transform of $\delta\Phi_{\boldsymbol{0}}(\boldsymbol{r})$ yields

$$\begin{aligned}\delta\Phi_{\boldsymbol{0}}(\boldsymbol{q}) &\equiv \frac{1}{A}\iint\delta\Phi(\boldsymbol{r})e^{i\boldsymbol{q}\cdot\boldsymbol{r}}d^2r\\ &= \frac{-en_s}{4\pi\varepsilon_s\kappa_b}\frac{2\pi}{qA}\int d\theta' r_0\delta r(\theta')e^{iqr_0\cos\theta'}.\end{aligned} \tag{B5}$$

The edge roughness function, $\delta r(\theta')$, is a periodic function with a periodicity of 2π. The Fourier series of $\delta r(\theta')$ is $\sum_n\Delta_n e^{in\theta'}$. Substituting $\delta r(\theta')$ with its Fourier series and incorporating $e^{iqr_0\cos\theta'} = \sum_m i^m J_m(qr_0)e^{im\theta'}$ in the above equation, we obtain

$$\begin{aligned}\delta\Phi_{\boldsymbol{0}}(\boldsymbol{q}) &= \frac{-en_s}{4\pi\varepsilon_s\kappa_b}\frac{2\pi}{qA}\sum_{nm}r_0\Delta_n i^m J_m(qr_0)\int d\theta' e^{in\theta'}e^{im\theta'}\\ &= \frac{-en_s}{\varepsilon_s\kappa_b}\frac{\pi}{qA}\left[r_0\Delta_0 J_0(qr_0) + 2\sum_{n=1}^{\infty}r_0\Delta_n i^n J_n(qr_0)\right].\end{aligned} \tag{B6}$$

In a similar way, it can be shown that the electric potential induced by the edge roughness of an antidot centered at $\boldsymbol{R}_0$

is $\delta\Phi_{\boldsymbol{R}_0}(\boldsymbol{q}) = e^{i\boldsymbol{q}\cdot\boldsymbol{R}_0}\delta\Phi_{\boldsymbol{0}}(\boldsymbol{q})$. However, assuming the same edge roughness for all the antidots results in an overestimation of the effective electric potential. To fix this, we substitute $\delta r(\theta')$ in Eq. (B5) with $\delta r(\theta' + \theta_r)$. θ_r is a random angle with uniform distribution in the $[-\pi, \pi]$ range. Now, we rewrite Eq. (B6) for an arbitrary θ_r:

$$\delta\Phi_{\boldsymbol{R}_0}(\boldsymbol{q}, \theta_r) = e^{i\boldsymbol{q}\cdot\boldsymbol{R}_0}\frac{-en_s}{\varepsilon_s\kappa_b}\frac{\pi}{qA} \times \left[r_0\Delta_0 J_0(qr_0) + 2\sum_{n=1}^{\infty} r_0\Delta_n i^n J_n(qr_0)\cos(n\theta_r)\right]. \tag{B7}$$

To obtain the effective electric potential induced by the ensemble of the antidots, we calculate the rms value of $\Phi_{\boldsymbol{R}_0}(\boldsymbol{q}, \theta_r)$:

$$\delta\Phi_{\boldsymbol{0}}(\boldsymbol{q}, \theta_r) = \frac{-en_s}{\varepsilon_s\kappa_b}\frac{\pi}{qA}\left[r_0\Delta_0 J_0(qr_0) + 2\sum_{n=1}^{\infty} r_0\Delta_n i^n J_n(qr_0)\cos(n\theta_r)\right]. \tag{B8}$$

Therefore, the effective electric potential is

$$\delta\Phi_{\text{eff}}(\boldsymbol{q}) = \sqrt{\sum_{\boldsymbol{R}_0}|e^{i\boldsymbol{q}\cdot\boldsymbol{R}_0}\delta\Phi_{\boldsymbol{0}}(\boldsymbol{q}, \theta_r(\boldsymbol{R}_0))|^2} = \sqrt{\frac{n_{AD}}{A}}\frac{e^2n_s}{\varepsilon_s\kappa_b}\frac{\pi}{q}\left[r_0\Delta_0 J_0(qr_0) + 2\sum_{n=1}^{\infty} r_0\Delta_n i^n J_n(qr_0)\right], \tag{B9}$$

where n_{AD} denotes the sheet density of antidots. Given the effective electric potential induced by the ensemble of the antidots, the transition rate from an eigenstate $|\boldsymbol{k}'l'\rangle$ to another eigenstate $|\boldsymbol{k}l\rangle$ via AER scattering is

$$\mathcal{S}(\boldsymbol{k}l, \boldsymbol{k}'l') = \frac{2\pi}{\hbar}\frac{n_{AD}}{A}\left[\frac{\pi e^2n_s}{\varepsilon_s\kappa_b|\varepsilon(\boldsymbol{k}-\boldsymbol{k}', \omega=0)|}\frac{1}{|\boldsymbol{k}-\boldsymbol{k}'|}\right]^2\left[r_0\Delta_0 J_0(qr_0) + 2\sum_{n=1}^{\infty} r_0\Delta_n i^n J_n(qr_0)\right]|(\boldsymbol{k}'l'|\boldsymbol{k}l)|^2\delta r(\epsilon_{\boldsymbol{k}l} - \epsilon_{\boldsymbol{k}'l'}). \tag{B10}$$

In graphene, $\epsilon_{\boldsymbol{k}l} = l\hbar v_F k$ and $|(\boldsymbol{k}'l'|\boldsymbol{k}l)|^2 = (1+\cos\alpha)/2$, with α being the angle between $\boldsymbol{k}$ and $\boldsymbol{k}'$. Assuming an exponential correlation function for the edge roughness function, $|\Delta_n|^2 = (\pi d/2)^{-1}\Delta^2\Lambda(1+\frac{4n^2}{d^2}\Lambda^2)^{-1}$, where Δ and Λ are the rms roughness and the correlation length, respectively. The expression for Δ_n is accurate for $e^{-\pi r_0/\Lambda} \ll 1$. The AER scattering for the exponentially correlated antidot edge roughness reads

$$\Gamma_m^{ab/em}(kl) = \frac{n_{AD}}{8\pi\hbar}\left(\frac{\pi e^2n_s}{\varepsilon_s\kappa_b\kappa^*}\right)^2\frac{k}{\hbar v_F}\int d\alpha\frac{\left[d\,\Delta_0 J_0\left(kd\sin\frac{\alpha}{2}\right)\right]^2 + 2\sum_{n=1}^{\infty}\left[d\,\Delta_n J_n\left(kd\sin\frac{\alpha}{2}\right)\right]^2}{\left(2k\sin\frac{\alpha}{2} + q_{\text{TF}}\right)^2}\frac{1-\cos^2\alpha}{2}, \tag{B11}$$

where q_{TF} represents the Thomas-Fermi screening wave vector [12].

[1] Z. Sun, A. Martinez, and F. Wang, Optical modulators with 2D layered materials, Nat. Photonics **10**, 227 (2016).

[2] Z. Sun, Electrically tuned nonlinearity, Nat. Photonics **12**, 383 (2018).

[3] C. Wang, M. Zhang, B. Stern, M. Lipson, and M. Lončar, Nanophotonic lithium niobate electro-optic modulators, Opt. Express **26**, 1547 (2018).

[4] A. Krasnok, M. Tymchenko, and A. Alu, Nonlinear metasurfaces: A paradigm shift in nonlinear optics, Mater. Today **21**, 8 (2018).

[5] D. Smirnova and Y. S. Kivshar, Multipolar nonlinear nanophotonics, Optica **3**, 1241 (2016).

[6] A. Grigorenko, M. Polini, and K. Novoselov, Graphene plasmonics, Nat. Photonics **6**, 749 (2012).

[7] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen *et al.*, Graphene plasmonics for tunable terahertz metamaterials, Nat. Nanotechnol. **6**, 630 (2011).

[8] F. H. Koppens, D. E. Chang, and F. J. García de Abajo, Graphene plasmonics: A platform for strong light-matter interactions, Nano Lett. **11**, 3370 (2011).

[9] 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, Polaritons in layered two-dimensional materials, Nat. Mater. **16**, 182 (2017).

[10] F. Karimi, A. H. Davoody, and I. Knezevic, Dielectric function and plasmons in graphene: A self-consistent-field calculation within a Markovian master equation formalism, Phys. Rev. B **93**, 205421 (2016).

[11] M. Jablan, H. Buljan, and M. Soljačić, Plasmonics in graphene at infrared frequencies, Phys. Rev. B **80**, 245435 (2009).

[12] E. H. Hwang and S. Das Sarma, Dielectric function, screening, and plasmons in two-dimensional graphene, Phys. Rev. B **75**, 205418 (2007).

[13] F. Karimi and I. Knezevic, Plasmons in graphene nanoribbons, Phys. Rev. B **96**, 125417 (2017).

[14] Z. Fang, Y. Wang, A. E. Schlather, Z. Liu, P. M. Ajayan, F. J. García de Abajo, P. Nordlander, X. Zhu, and N. J. Halas, Active tunable absorption enhancement with graphene nanodisk arrays, Nano Lett. **14**, 299 (2014).

[15] H. Liu, Y. Liu, and D. Zhu, Chemical doping of graphene, J. Mater. Chem. **21**, 3335 (2011).

[16] D. K. Efetov and P. Kim, Controlling Electron-Phonon Interactions in Graphene at Ultrahigh Carrier Densities, Phys. Rev. Lett. **105**, 256805 (2010).

[17] F. Karimi, A. H. Davoody, and I. Knezevic, Nonlinear optical response in graphene nanoribbons: The critical role of electron scattering, Phys. Rev. B **97**, 245403 (2018).
[18] J. D. Cox, I. Silveiro, and F. J. García de Abajo, Quantum effects in the nonlinear response of graphene plasmons, ACS Nano **10**, 1995 (2016).
[19] K. Ooi and D. Tan, Nonlinear graphene plasmonics, Proc. R. Soc. A. **473**, 20170433 (2017).
[20] Y. Ang, Q. Chen, and C. Zhang, Nonlinear optical response of graphene in terahertz and near-infrared frequency regime, Front. Optoelectron. **8**, 3 (2015).
[21] Q. Feng, H. Cong, B. Zhang, W. Wei, Y. Liang, S. Fang, T. Wang, and J. Zhang, Enhanced optical Kerr nonlinearity of graphene/Si hybrid waveguide, Appl. Phys. Lett. **114**, 071104 (2019).
[22] X. Sun, B. Zhang, Y. Li, X. Luo, G. Li, Y. Chen, C. Zhang, and J. He, Tunable ultrafast nonlinear optical properties of graphene/MoS_2 van der Waals heterostructures and their application in solid-state bulk lasers, ACS Nano **12**, 11376 (2018).
[23] S. Yu, X. Wu, Y. Wang, X. Guo, and L. Tong, 2D materials for optical modulation: Challenges and opportunities, Adv. Mater. **29**, 1606128 (2017).
[24] S.-Y. Hong, J. I. Dadap, N. Petrone, P.-C. Yeh, J. Hone, and R. M. Osgood, Jr., Optical Third-Harmonic Generation in Graphene, Phys. Rev. X **3**, 021014 (2013).
[25] T. Gu, N. Petrone, J. McMillan, A. Van Der Zande, M. Yu, G. Lo, D. Kwong, J. Hone, and C. Wong, Regenerative oscillation and four-wave mixing in graphene optoelectronics, Nat. Photonics **6**, 554 (2012).
[26] N. Vermeulen, D. Castello-Lurbe, M. Khoder, I. Pasternak, A. Krajewska, T. Ciuk, W. Strupinski, J. Cheng, H. Thienpont, and J. V. Erps, Graphene's nonlinear-optical physics revealed through exponentially growing self-phase modulation, Nat. Commun. **9**, 2675 (2018).
[27] K. Alexander, N. A. Savostianova, S. A. Mikhailov, B. Kuyken, and D. V. Thourhout, Electrically tunable optical nonlinearities in graphene-covered SiN waveguides characterized by four-wave mixing, ACS Photonics **4**, 3039 (2017).
[28] M. Kauranen and A. V. Zayats, Nonlinear plasmonics, Nat. Photonics **6**, 737 (2012).
[29] J. D. Cox and F. J. G. De Abajo, Electrically tunable nonlinear plasmonics in graphene nanoislands, Nat. Commun. **5**, 5725 (2014).
[30] J. Bai, X. Zhong, S. Jiang, Y. Huang, and X. Duan, Graphene nanomesh, Nat. Nanotechnol. **5**, 190 (2010).
[31] Z. Zeng, X. Huang, Z. Yin, H. Li, Y. Chen, H. Li, Q. Zhang, J. Ma, F. Boey, and H. Zhang, Fabrication of graphene nanomesh by using an anodic aluminum oxide membrane as a template, Adv. Mater. **24**, 4138 (2012).
[32] K. Y. Yeung, J. Chee, H. Yoon, Y. Song, J. Kong, and D. Ham, Far-infrared graphene plasmonic crystals for plasmonic band engineering, Nano Lett. **14**, 2479 (2014).
[33] X. Zhu, W. Wang, W. Yan, M. B. Larsen, P. Bøggild, T. G. Pedersen, S. Xiao, J. Zi, and N. A. Mortensen, Plasmon–phonon coupling in large-area graphene dot and antidot arrays fabricated by nanosphere lithography, Nano Lett. **14**, 2907 (2014).
[34] W. Gao, J. Shu, C. Qiu, and Q. Xu, Excitation of plasmonic waves in graphene by guided-mode resonances, ACS Nano **6**, 7806 (2012).
[35] P. Q. Liu, F. Valmorra, C. Maissen, and J. Faist, Electrically tunable graphene anti-dot array terahertz plasmonic crystals exhibiting multi-band resonances, Optica **2**, 135 (2015).
[36] A. Y. Nikitin, F. Guinea, and L. Martin-Moreno, Resonant plasmonic effects in periodic graphene antidot arrays, Appl. Phys. Lett. **101**, 151119 (2012).
[37] B. Zhu, G. Ren, Y. Gao, Y. Yang, Y. Lian, and S. Jian, Graphene-coated tapered nanowire infrared probe: A comparison with metal-coated probes, Opt. Express **22**, 24096 (2014).
[38] L. Wu, Y. Dong, J. Zhao, D. Ma, W. Huang, Y. Zhang, Y. Wang, X. Jiang, Y. Xiang, J. Li, Y. Feng, J. Xu, and H. Zhang, Kerr nonlinearity in 2D graphdiyne for passive photonic diodes, Adv. Mater. **31**, 1807981 (2019).
[39] Z. Wang, Y. Xu, S. C. Dhanabalan, J. Sophia, C. Zhao, C. Xu, Y. Xiang, J. Li, and H. Zhang, Black phosphorus quantum dots as an efficient saturable absorber for bound soliton operation in an erbium doped fiber laser, IEEE Photonics J. **8**, 1 (2016).
[40] Z. Xie, F. Zhang, Z. Liang, T. Fan, Z. Li, X. Jiang, H. Chen, J. Li, and H. Zhang, Revealing of the ultrafast third-order nonlinear optical response and enabled photonic application in two-dimensional tin sulfide, Photon. Res. **7**, 494 (2019).
[41] L. Wu, W. Huang, Y. Wang, J. Zhao, D. Ma, Y. Xiang, J. Li, J. S. Ponraj, S. C. Dhanabalan, and H. Zhang, 2D tellurium based high-performance all-optical nonlinear photonic devices, Adv. Funct. Mater. **29**, 1806346 (2019).
[42] R. W. Boyd, *Nonlinear Optics*, 3rd ed. (Academic Press, Burlington, MA, 2008).
[43] F. Karimi, S. Soleimanikahnoj, and I. Knezevic, Tunable plasmon-enhanced second-order optical nonlinearity in transition metal dichalcogenide nanotriangles, Phys. Rev. B **103**, L161401 (2021).
[44] M. A. Yamoah, W. Yang, E. Pop, and D. Goldhaber-Gordon, High-velocity saturation in graphene encapsulated by hexagonal boron nitride, ACS Nano **11**, 9914 (2017).
[45] N. Petrone, I. Meric, T. Chari, K. L. Shepard, and J. Hone, Graphene field-effect transistors for radio-frequency flexible electronics, IEEE J. Electron Devices Soc. **3**, 44 (2015).
[46] N. Petrone, T. Chari, I. Meric, L. Wang, K. L. Shepard, and J. Hone, Flexible graphene field-effect transistors encapsulated in hexagonal boron nitride, ACS Nano **9**, 8953 (2015).
[47] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.108.035414 for more details on the role of different substrates and the real and imaginary parts of susceptibility.