Resonant Plasmonic Terahertz Detection in Gated Graphene *p-i-n* Field-Effect Structures Enabled by Nonlinearity from Zener-Klein Tunneling

V. Ryzhii,[1,*] T. Otsuji,[1] M. Ryzhii,[2] V. Mitin,[3] and M. S. Shur[4]

[1] *Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan*
[2] *Department of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu 965-8580, Japan*
[3] *Department of Electrical Engineering, University at Buffalo, SUNY, Buffalo, New York 14260, USA*
[4] *Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, USA*

(Received 22 June 2022; revised 28 July 2022; accepted 4 August 2022; published 9 September 2022)

We show that resonant plasmonic detection dramatically increases the sensitivity of the terahertz detectors based on a gated graphene *p-i-n* (GPIN) field-effect transistor (FET) structure. In the proposed device, the gated *p* and *n* regions serve as the hole and electron reservoirs and the terahertz resonant plasma cavities. The current-voltage (*I-V*) characteristics are strongly nonlinear due to the Zener-Klein interband tunneling in the reverse-biased *i* region between the gates. The terahertz signal rectification by this region enables the terahertz detection. The resonant excitation of the hole and electron plasmonic oscillations results in a substantial increase in the terahertz detector responsivity at the signal frequency close to the plasma frequency and its harmonics. Because of the transit-time effects, the GPIN-FET response at the higher plasmonic modes could be stronger than for the fundamental mode. Our estimates predict the detector responsivity up to a few 10^5 V/W at room temperature, much larger than for other electronic terahertz detectors, such as Schottky diodes, *p-n* junctions, Si CMOS, and III-V and III-N HEMTs.

DOI: 10.1103/PhysRevApplied.18.034022

I. INTRODUCTION

Short channel field-effect transistors (FETs) can serve as effective terahertz detectors [1]. Such detectors could operate in a resonant regime when the detection is strongly enhanced by plasmonic resonances or in the rectification regime when the plasmonic oscillations are overdamped. In either case, the detector responsivity is proportional to the nonlinearity of the current-voltage (*I-V*) characteristics. Different nonlinearity mechanisms enable the terahertz signal rectification in plasmonic FET detectors, including the hydrodynamic nonlinearity and the barrier rectification in the Schottky junctions, *p-n* junctions, or electrostatic barriers (see, for example, [2–4]). The FET terahertz detectors with the graphene layer (GL) channel can exhibit markedly enhanced performance [5–12] due to the unique electron and hole transport properties of GLs [13–16], in particular, high electron (hole) mobility and directed velocity at elevated temperatures. The specific features of the GL band structure enable the Zener-Klein interband tunneling [17–22] leading to a very strong nonlinearity that can be used for the rectification and detection of the terahertz signals. In this paper, we evaluate the proposed terahertz detector based on a lateral graphene *p-i-n* FET (GPIN-FET) detector structure with the gated *p* and *n* regions of the GL channel. This device combines the advantages of the strong *I-V* nonlinearity and plasmonic resonant detection enhancement enabled by high carrier mobility in GLs and transit-time effects. Such a combination of high nonlinearity and plasmonic effects leads to a remarkably high responsivity. Our estimates predict the detector responsivity up to a few 10^5 V/W, markedly exceeding that of Schottky diodes, *p-n* junctions, Si CMOS, and III-V and III-N HEMTs.

II. MODEL

We consider the GPIN-FET detector structure based on a GL channel embedded in the dielectric [for example, hexagonal boron nitride (*h*-BN)]. The channel is covered by two highly conducting gates. The gate voltages of different polarity, $\pm V_g$, are applied to these gates. As a result, the channel comprises the undoped *i*regions between the gates (of the length $2l$) and the electrically doped *p* and *n* regions beneath the gates (of the length L). The lengths of the *p* and *n* regions are close to the length of the gates (L is somewhat larger than the gate length due to the finite screening length or the gate fringe effect). The channel is bounded by the source and drain contacts, between which

*v-ryzhii@riec.tohoku.ac.jp

 © 2022 American Physical Society

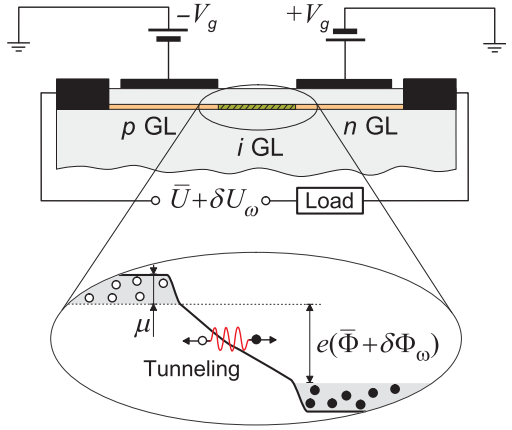

FIG. 1. Schematic view of the cross sections of the GPIN-FET structure cross section with gated electrostatically doped p and n regions in the GPIN-FET channel. Inset shows the potential profile in the channel.

the bias voltage $\overline{U}$ is applied. Figure 1 shows the GPIN-FET structure with the electrically disconnected gates and the potential profile in the GPIN-FET channel at the dc condition (inset) when the bias voltage is applied between the side (source/drain) contacts.

In the present paper we consider the GPIN-FET structures, where the Coulomb drag effect in the gated regions [23–28] is weak. This corresponds to GPIN-FET structures, in which more liberal and rather practical conditions are fulfilled: stronger scattering on impurities, defects, and acoustic phonons (with the collision frequencies $\nu \gtrsim 1\ \mathrm{ps}^{-1}$), and relatively long gated regions ($L \gtrsim 1\ \mu$m). The i region is assumed to be fairly short ($2l = 0.1$–$0.2\ \mu$m), so that the transport of the holes and electrons generated in this region due to the Zener-Klein tunneling is ballistic [29–33]. Considering that the i region is depleted under the operational conditions, the conditions of the ballistic transport in this region with the above length can be realized even at room temperature [34] and, naturally, at lowered temperatures [35]. The holes and electrons generated due to tunneling are directed primarily along the electric field in the i region, i.e., in the x direction (the in-plane direction along the GL channel) from the p and n regions, and propagate ballistically with the velocities, $v = \pm v_W$.

Apart from the dc bias voltage $\overline{U}$, an ac signal voltage $\delta U_\omega \exp(-i\omega t)$ is applied, where δU_ω and ω are the amplitude and frequency of the incoming terahertz signal. Thus, both the dc and ac voltage drops across the GPIN-FET structure and are equal to $\overline{V} = \overline{U} - \overline{U}^{\mathrm{load}}$ and $\delta V_\omega = \delta U_\omega - \delta U_\omega^{\mathrm{load}}$, where $\overline{U}^{\mathrm{load}}$ and $\delta U_\omega^{\mathrm{load}}$ are the dc and ac components of the load voltage. These components depend on the load impedance Z^{load}. The latter can be presented as $Z^{\mathrm{load}} = [1/r^{\mathrm{load}} - i\omega C^{\mathrm{load}}]^{-1}$, where r^{load} and C^{load} are the load resistance and capacitance, respectively. We assume that C^{load} is sufficiently large, so that the ac voltage drop across it is negligibly small, therefore, $\delta U_\omega^{\mathrm{load}} \simeq 0$ and $\delta V_\omega = \delta U_\omega$. The real part of the load impedance, i.e., its resistance, r^{load} determines the dc output signal.

III. LINEAR RESPONSE

When both the dc voltage $\overline{V}$ and the periodic signal voltage $\delta V_\omega \exp(-i\omega t)$ drop across the GPIN FET intrinsic part [$V = \overline{V} + \delta V_\omega \exp(-i\omega t)$], the carrier current density in the i region (and in other sections of the channel) is equal to $J^i = \overline{J}^i + \Delta J^i$. Here $\overline{J}^i$ is the dc component. The component ΔJ^i comprises the linear ac component δJ^i_ω proportional to δV_ω and the ac rectified component ΔJ^i_ω. The latter is due to the nonlinear I-V characteristics of the i region at the conditions of the Zener-Klein tunneling.

The voltage across the devices is distributed between the i region and the gated regions.

A. Dynamic conductance of the *i* region

Considering the bladelike configuration of the conducting areas surrounding the i region, the pertinent spatial electric-field distribution in this region [36–39], and using the general formulas for the Zener-Klein tunneling probability in graphene [17–19], one can arrive at the following expression for the dc current density per unit width in the direction along the gate edges [22]:

$$\overline{J} = a\frac{ev_W}{\sqrt{2l}}\left(\frac{e\overline{\Phi}}{\hbar\, v_W}\right)^{3/2}, \tag{1}$$

where a is a numerical parameter. At moderate $\overline{\Phi}$ when the carrier space charge in the i regions is weak,

$$a = \frac{\Gamma(1/4)\Gamma(1/2)}{\Gamma(3/4)}\frac{1}{2\pi^{7/2}},$$

where $\Gamma(z)$ is the gamma function. Equation (1) corresponds to the i-region linear dynamic ac conductance (see Appendix A)

$$\sigma^i_\omega = \sigma^i F_\omega - i\omega\, c^i = \sigma^i(F_\omega - i\omega\,\tau^i). \tag{2}$$

Here, $\sigma^i = d\overline{J}^i/d\overline{\Phi}$ is the i-region dc differential conductance, $\tau^i = c^i/\sigma^i$ is the i-region recharging time, and c^i is the geometrical capacitance per unit width of the device determined mainly by the dielectric constant κ of the isolating material surrounding the GL, and geometrical parameters (see, for example, [36–39]). If the GL is deeply embedded into the isolating material $c^i \propto \kappa$. In the case of a free GL top surface between the gates, $c^i \propto \kappa^{eff} = (\kappa + 1)/2$. A thin passivation layer can also affect c^i.

According to Eq. (1),

$$\sigma^i = b\frac{e^2}{\hbar}\sqrt{\frac{e\overline{\Phi}}{2l\hbar\, v_W}}, \tag{3}$$

where $b = 3a/2 \simeq 0.0716$. The quantity

$$F_\omega = \mathcal{J}_0(\omega\, t^i/2)\, e^{i\omega\, t^i/2} \tag{4}$$

reflects the signal frequency dependence of the i-region dynamic conductivity determined by the ac current induced by the carriers propagating between the p and n regions (and the pertinent gates), particularly, by the finiteness of their transit time $t^i = 2l/v_W$, where $\mathcal{J}_0(s)$ is the Bessel function of the first kind.

Using Eqs. (2) and (4), we obtain the following expressions for the real and imaginary parts of σ^i_ω:

$$\mathrm{Re}\,\sigma^i = \sigma^i \mathcal{J}_0(\omega\, t^i/2)\, \cos(\omega\, t^i/2), \tag{5}$$

$$\mathrm{Im}\,\sigma^i = \sigma^i[\mathcal{J}_0(\omega\, t^i/2)\, \sin(\omega\, t^i/2) - \omega\tau^i]. \tag{6}$$

The quantity $\mathrm{Re}\,\sigma^i$ can be both positive and negative. However, at $\omega\, t^i < \pi$, $\mathrm{Re}\,\sigma^i > 0$. In the range $\omega\, t^i \lesssim 4.8$, where $\sin(\omega\, t^i/2) > 0$, the first term on the right-hand side of Eq. (6) is positive. The latter implies that this term corresponds to the kinetic inductance of the holes and electrons in the i region. In certain ranges of elevated frequencies, the product $\mathcal{J}_0(\omega\, t^i/2)\sin(\omega\, t^i/2)$ can be positive, although both factors are negative.

For $2l = 0.2$ μm and $\overline{\Phi} = 100 - 200$ mV, Eqs. (1) and (4) yield $\overline{J}^i \simeq 0.34$–$0.95$ A/cm and σ^i in the range from 520–733 S/m (from 4.67–6.59 ps^{-1}). In the case of a sufficiently short i region the ballistic transport takes place even at higher voltages $\overline{\Phi} > \hbar\omega_0/e$ despite the spontaneous emission of optical phonons, and the current density through this region and, hence, the differential conductance can substantially exceed the above estimates.

Considering Eq. (2), we arrive at the following equation for the linear ac current density in the i region expressed via the ac potential drop $\delta\Phi_\omega = (\delta\varphi^i_\omega|_{x=l} - \delta\varphi^i_\omega|_{x=-l})$ across this region:

$$\delta J_\omega = \sigma^i_\omega \delta\Phi_\omega. \tag{7}$$

B. Plasmonic response of the gated regions

To express $(\delta\varphi^i_\omega|_{x=l} - \delta\varphi^i_\omega|_{x=-l})$ via δV_ω, we find the spatial distributions of the ac potential in the gated region accounting for its nonuniformity associated with the plasmonic effects.

For the densities of the ac current in the p and n region ($-L - l < x < -l$ and $l < xl + L$), we have

$$\delta J^g_\omega = -\sigma^g_\omega L\frac{d\delta\varphi_\omega}{dx}\bigg|_{x=\pm l}, \tag{8}$$

where $\sigma^g_\omega = \sigma^g[i\nu/(\omega + i\nu)]$ and $\sigma^g = (e^2\mu/\pi\hbar^2\nu\, L)$ are the p and n regions (gated) ac and dc conductances, respectively, ν is the frequency of hole and electron momentum relaxation on impurities, acoustic phonons, and $\delta\varphi_\omega = \delta\varphi_\omega(x,y)|_{y=0}$ expresses the ac potential spatial distribution along the x axis directed in the GPIN-FET-channel plane ($y = 0$). The frequency dependence given by Eq. (8) accounts for the kinetic inductance of the gated regions.

For the gated sections of the channel [see Fig. 1(a)], we solve the linearized hydrodynamic equations for the hole and electron plasmas in the related p and n regions, disregarding the nonuniformity of the dc potential and carrier density distributions, and arrive at the following equation for the spatial distribution of the ac potential $\delta\varphi_\omega$ accounting for the plasmonic response of the gated regions (see [12] and the references therein):

$$\frac{d^2\delta\varphi_\omega}{dx^2} + \frac{\omega(\omega + i\nu)}{s^2}\left(\delta\varphi_\omega \mp \frac{\eta}{2}\delta V_\omega\right) = 0. \tag{9}$$

Here the upper/lower sign is related to the n region/p region, $s = \sqrt{4e^2\mu\, w/\kappa\hbar^2}$ is the plasma velocity, and $\mu = \mu^p = \mu^n$ is the carrier Fermi energy in the gated regions of both types.

The quantity $\eta = C_{cg}/(C_{cg} + C_g)$ characterizes the contact-gate coupling, where $C_{cg} \propto \kappa$ [or $C_{cg} \propto \kappa^{eff} = (\kappa + 1)/2$] and $C_g = (\kappa\, L/4\pi\, w)$ being the pertinent capacitances for the bladelike contacts and gates, where w is the gate layer thickness. Normally, $C_{cg} \ll C_g$, hence $\eta \ll 1$.

Equation (9) governs the ac potential in the p region ($-L - l < x < -l$) and in the n region ($l < x < l + L$). It accounts for both the gate and contact-gate capacitances and the kinetic inductance of the holes and electrons in the gated regions.

The boundary conditions at the edges of the gated regions are

$$\delta\varphi_\omega|_{\pm(l+L)} = \pm\frac{1}{2}\delta V_\omega, \tag{10}$$

$$\sigma^i_\omega(\delta\varphi_\omega|_{x=l} - \delta\varphi_\omega|_{x=-l}) = \sigma^g_\omega L\frac{d\delta\varphi_\omega}{dx}\bigg|_{x=\pm l}. \tag{11}$$

Using the solution of Eqs. (9)–(11) [see Appendix B, Eqs. (B11) and (B12)] and considering that

$\delta\Phi_\omega = (\delta\varphi_\omega|_{x=l} - \delta\varphi_\omega|_{x=-l})$, we obtain

$$\delta\Phi_\omega = P_\omega \delta V_\omega, \qquad \delta J_\omega = \sigma^i_\omega P_\omega \delta V_\omega, \tag{12}$$

with $\delta V_\omega = \delta U_\omega$. Here

$$P_\omega = \frac{\eta\cos(æ_\omega L) + 1 - \eta}{\cos(æ_\omega L) + (1/\xi_\omega)(\sin(æ_\omega L)/(æ_\omega L))}. \tag{13}$$

Here the following notations have been introduced: $æ_\omega = \pi\sqrt{\omega(\omega + i\nu)}/2\Omega L$, $\xi_\omega = \sigma^g_\omega/2\sigma^i_\omega = (\sigma^g/2\sigma^i)[i\nu/(\omega + i\nu)(F_\omega - i\omega\tau^i)]$. The quantity

$$\Omega = \frac{\pi e}{\hbar L}\sqrt{\frac{\mu w}{\kappa}} \tag{14}$$

is the plasma frequency of the gated regions (see, for example, [12,40–43]).

Equation (14) accounts for the fact that the electron liquid in the gated portion of the channel at the carrier densities under consideration is degenerate.

Considering that in the real case $\eta \ll 1$, the function P_ω is weakly sensitive to the contact-gate capacitive coupling, in particular, to the parasitic capacitance C_{cg} (see Sec. VII), and Eq. (13) can be somewhat simplified as

$$P_\omega \simeq \frac{1}{\cos(æ_\omega L) + (1/\xi_\omega)(\sin(æ_\omega L)/(æ_\omega L))}. \tag{15}$$

IV. RECTIFIED CURRENT AND PLASMONIC FACTOR

The I-V characteristic given by Eq. (1) corresponds to the nonlinear current density component, which comprises the rectified component and the ac current harmonics,

$$\Delta\tilde{J}_\omega = \beta\,\delta\Phi^2_\omega, \tag{16}$$

where

$$\beta = \frac{1}{2}\frac{d^2\overline{J}}{d\,\overline{\Phi}^2} = \frac{\sigma^i}{4\overline{\Phi}} = \frac{be^{5/2}}{4\sqrt{2l\hbar^3 v_W \overline{\Phi}}} \propto \frac{1}{\sqrt{\overline{\Phi}}} \tag{17}$$

is the parameter characterizing the nonlinearity of the i-region tunneling I-V characteristics given by Eq. (1), i.e., the curvature of these characteristic.

According to the Kirchhoff law,

$$\overline{\Phi} = \frac{\overline{U}}{\gamma}, \qquad \delta\Phi_\omega = \delta U_\omega, \tag{18}$$

where $\gamma = 1 + 2H\sigma^i r^{\rm load}/3$. Here we accounted for the fact that the i-region dc conductance is equal to $2\sigma^i/3$.

Equation (16) yields the following expression for the density of the rectified current $\Delta\overline{J}_\omega = (\omega/2\pi)\int_0^{2\pi/\omega} dt\Delta\tilde{J}_\omega$:

$$\Delta\overline{J}_\omega = \delta V^2_\omega \frac{\sigma^i\gamma}{8\overline{U}}|P_\omega|^2\delta U^2_\omega. \tag{19}$$

As seen from Eq. (19), the rectified current density $\Delta\overline{J}_\omega$, as will be seen in the following, and the frequency dependence of the GPIN-FET detector responsivity, are determined by the plasmonic factor $|P_\omega|^2$. According to Eq. (16), this factor is given by

$$|P_\omega|^2 = \left|\frac{\eta\cos(æ_\omega L) + 1 - \eta}{\cos(æ_\omega L) + (1/\xi_\omega)(\sin(æ_\omega L)/(æ_\omega L))}\right|^2. \tag{20}$$

Figure 2 shows the frequency dependences of the plasmonic factor $|P_\omega|^2$ calculated for the GPIN FETs with $L = 1.0$ μm, $w = 50$ nm, $\mu = 100$ meV, $\kappa = 4.5$,

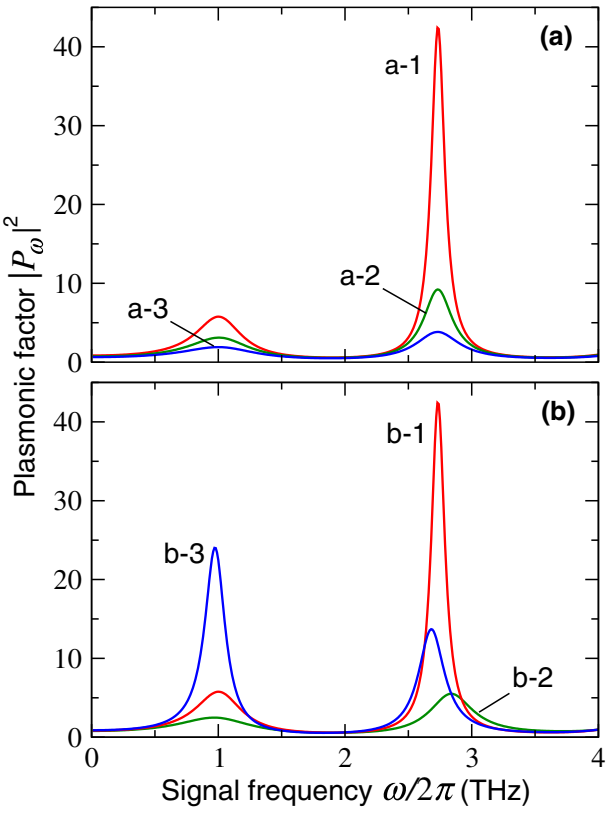

FIG. 2. Plasmonic factor $|P_\omega|^2$ versus signal frequency $f = \omega/2\pi$ (a) for GPIN FETs (a-1)–(a-3) with different values of frequency, ν, of the carrier momentum relaxation in the gated regions and, consequently, different conductivity, σ^g, of these regions; and (b) for GPIN FETs (b-1)–(b-3) with different lengths of the i region $2l$, corresponding to different differential conductivity σ^i, and transit time t^i ($\overline{U} = 200$ mV). The parameters of the samples are given in Table I.

TABLE I. GPIN-FET parameters.

Sample	$2l$ (μm)	t^i (ps)	τ^i (ps)	L (μm)	w (nm)	κ	μ (meV)	ν (ps^{-1})	$\sigma^g/2\sigma^i$	$\Omega/2\pi$ (THz)
a-1[a]	0.2	0.2	0.1	1.0	15	4.5	100	1.0	12	1.0
a-2	0.2	0.2	0.1	1.0	15	4.5	100	2.0	6	1.0
a-3	0.2	0.2	0.1	1.0	15	4.5	100	3.0	4	1.0
b-1[a]	0.2	0.2	0.1	1.0	15	4.5	100	1.0	12	1.0
b-2	0.1	0.1	0.1	1.0	15	4.5	100	1.0	8.5	1.0
b-3	0.4	0.4	0.1	1.0	15	4.5	100	1.0	16.9	1.0
c-1[a]	0.2	0.2	0.1	1.0	15	4.5	100	1.0	12	1.0
c-2	0.2	0.2	0.1	1.0	15	4.5	144	1.0	14.4	1.2
c-3	0.2	0.2	0.1	1.0	15	4.5	64	1.0	9.6	0.8
d-1[a]	0.2	0.2	0.1	1.0	15	4.5	100	1.0	12	1.0
d-2	0.2	0.2	0.2	1.0	33	10.0	100	1.0	12	1.0
d-3	0.2	0.2	0.3	1.0	50	15.0	100	1.0	12	1.0

[a]Samples with identical parameters.

$\tau^i \simeq 0.1$ ps^{-1}($c^i \simeq 0.5$), $\Omega/2\pi = 1.0$ THz, and $\eta = 0.1$ at $\overline{\Phi} = 100$ mV, i.e., $\overline{U} \simeq 200$ mV and $\gamma \simeq 2$ (see Table I).

As seen in Fig. 2, $|P_\omega|^2$ and, consequently, the rectified current components given by Eq. (19), exhibit two pronounced maxima (in the frequency range under consideration, $\omega/2\pi \leq 4$ THz).

The first maximum at $\omega/2\pi \simeq 1$ THz corresponds to the signal frequency close to the gated regions plasma frequency $\Omega/2\pi$. It is obviously related to the excitation (by the incoming signal) of the fundamental mode of a standing plasma wave with the wave number $q_1 \simeq \pi/2L$ and the maximum amplitude at $x = \pm l$ having the opposite phases. As a result, $|\delta\Phi_\omega|$ and $|\delta J_\omega|$ are maximal. The collisional damping of these plasma oscillations results in lowering of the plasma resonance peaks [see the curves corresponding to different ν in Fig. 2(a)].

The second resonant peaks of $|P_\omega|^2$ at $\omega/2\pi \lesssim 3$ THz seen in Fig. 2 correspond to the excitation of the plasmonic mode with the wave number $q_3 \simeq 3\pi/2L$. It is instructive that the peaks associated with this mode can be higher than those related to the fundamental mode.

At elevated signal frequencies Eqs. (19) and (20) yield

$$\Delta\overline{J}_\omega \propto |P_\omega|^2 \propto \left(\frac{\overline{\omega}}{\omega}\right)^2, \tag{21}$$

where

$$\overline{\omega} = \frac{1}{\tau^i}\left(\frac{\pi\nu}{2\Omega}\right)\left(\frac{\sigma^g}{2\sigma^i}\right).$$

For the parameters of samples (a-1)–(a-3), we obtain $\overline{\omega}/2\pi \simeq 4.8$ THz. The roll-off of $\Delta\overline{J}_\omega$ and $|P_\omega|^2$ with increasing frequency is determined by the charging time τ^i, i.e., is associated with the i-region geometrical capacitance c^i. Considering that $\sigma^g \propto 1/\nu$, one can find that $\overline{\omega}$ is independent of ν.

V. GPIN-FET DETECTOR RESPONSIVITY

The current responsivity (ampere-watt) of GPIN FETs operating as the terahertz detectors is given by

$$R^J_\omega = \frac{\Delta\overline{J}_\omega H}{SI_\omega}, \tag{22}$$

where H is the lateral size of the GPIN FET in the direction perpendicular to the gate edges, $S = \lambda^2_\omega g/4\pi$ and $g \sim 1.5$ are the antenna aperture and gain, $\lambda_\omega = 2\pi c/\omega$ and I_ω are the wavelength and intensity of the incoming radiation, and c is the speed of light in vacuum. Considering that $\delta U^2_\omega = 8\pi\lambda^2_\omega I_\omega/c$, using Eq. (22), we obtain

$$R^J_\omega = \overline{R}^J|P_\omega|^2, \qquad \overline{R}^J = \frac{4\pi^2 H\sigma^i}{cg}\frac{\gamma}{\overline{U}}. \tag{23}$$

For the voltage (volt-watt) responsivity $R^V_\omega = r^{\rm load}R^J_\omega$, assuming that the load resistance is optimized ($r^{\rm load} = 3/2\sigma^i H$, i.e., $\gamma \simeq 2$ if $\sigma^i \ll \sigma^g/2$), we arrive at the following universal formula:

$$R^V_\omega = \overline{R}^V|P_\omega|^2, \qquad \overline{R}^V = \frac{6\pi^2\gamma}{cg\overline{U}}. \tag{24}$$

For $\overline{U} = 200$–400 mV, we obtain $\overline{R}^V \simeq (1.12$–$0.56)\times10^4$ V/W. If $2l = 0.2$ μm, $H = 10$ μm, and $\overline{U} = 200$–400 mV, the optimized load resistance is equal to $r^{\rm load} \simeq 307$–216 Ω. Naturally, the plasmonic resonance may lead to $R^V_\omega \gg \overline{R}^V$.

A decrease in $\overline{R}^V$ and, hence, in R^V_ω with the rise of the bias dc voltage $\overline{U}$ is attributed to a decrease in the current-voltage characteristic nonlinearity parameter β [see Eq. (17)], and to a decrease in the channel resistance with rising bias voltage $\overline{U}$. The latter requires the pertinent decrease in the optimized load resistance $r^{\rm load}$. However,

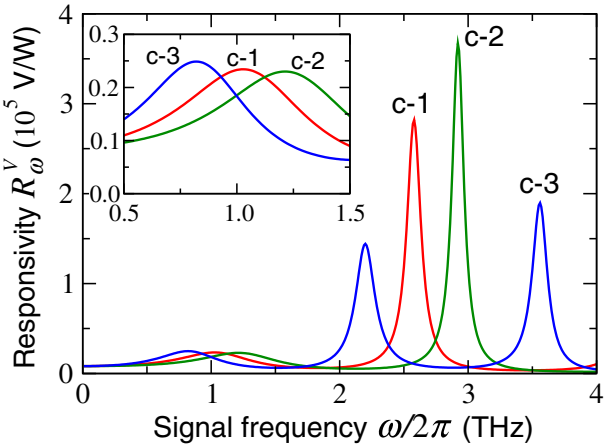

FIG. 3. Frequency dependences of Volt-Watt responsivity for the GPIN FETs (c-1)–(c-3) with different values of the carrier Fermi energy μ, conductance σ^g, and plasma frequency Ω ($\overline{U} =$ 200 mV). Inset shows the vicinity of the fundamental resonance in more detail.

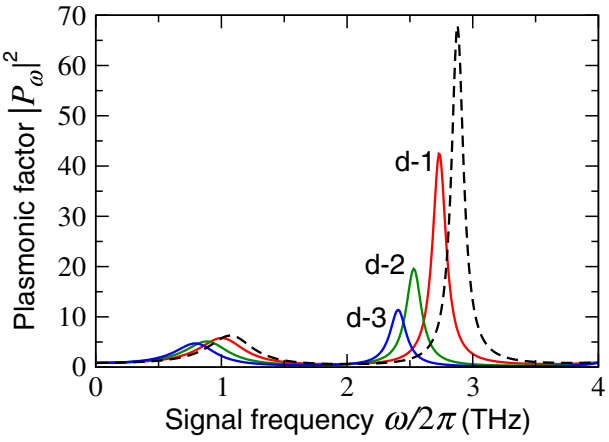

FIG. 4. Plasmonic factor $|P_\omega|^2$ versus signal frequency $f = \omega/2\pi$ for GPIN FETs (d-1)–(d-3), which differ by charging time τ^i (proportional to gate layer dielectric constants κ). The dashed line corresponds to $\tau^i = 0.05$ ps ($\kappa = 2.25$).

lowering of the bias voltage is limited by the thermionic and thermogeneration processes in the *i* region.

Figure 3 shows the GPIN-FET responsivity R_ω^V as a function of the signal frequency $\omega/2\pi$ calculated using Eqs. (23) and (24) involving Eq. (20) for the devices (c-1)–(c-3) with the parameters presented in Table I assuming that $\overline{U} = 200$ mV.

First of all, one can see that GPIN FETs can reveal fairly high peak responsivity. Second, the responsivity peaks corresponding to the higher plasma modes can be markedly higher than the fundamental peak (see the discussion in Sec. VI). Third, the second peaks c-1 and c-2 (in the main plot and the inset in Fig. 3) are positioned at the frequencies $\omega/2\pi$ somewhat lower than $\Omega/2\pi = 1.0$ THz and $\Omega/2\pi = 1.2$ THz, respectively. As mentioned in Sec. IV, this is due to the effect of the *i*-region capacitance on the resonant frequency deviating it from the plasma frequency of the gated region Ω. The same is valid for the second and third peaks (c-3) in the main plot and the inset.

VI. DISCUSSION (ANALYSIS)

A. Temperature dependence and heights of the resonant peaks

The obtained formulas for the GPIN-FET responsivity R_ω^V do not explicitly account for the temperature dependence, at least at $\mu \gg T$, i.e., in the situations under consideration. This is because of the tunneling nature of the current in the GPIN FETs. However, the plasmonic resonant factor, which determines the maximum values of R_ω^V is sensitive to the collision frequency ν. The latter is usually smaller at lower temperatures due to a decrease in the carrier momentum relaxation on acoustic phonons. Hence, lowering of the temperature can result in a marked sharpening of the resonant responsivity peaks and promote a substantial increase in the resonant responsivity.

Figures 2 and 3 show some deviation of the resonant peaks position from the exact plasmonic resonances $\omega/2\pi = \Omega/2\pi = 1$ THz and $\omega/2\pi = 3$ THz—the peaks are shifted toward smaller frequencies. This is attributed to the collisional damping and to the contribution of the *i*-region geometrical capacitance. Indeed, as seen from Fig. 4, the GPIN FETs with longer charging time τ^i, i.e., a larger *i*-region geometrical capacitance c^i due to a larger dielectric constant of the gate layer κ (and the same collisional frequency ν), exhibit smaller resonant frequencies. The gate layer in the samples (a-1)–(d-1) is assumed to be made of *h*-BN ($\kappa = 4.5$), whereas the gate layers in the samples d-2 and d-3 are made of SiC ($\kappa = 10$) and HfO_2 ($\kappa = 15$), respectively.

Table I lists the values of κ for the samples under consideration. The assumed value of κ for GPIN FET passivated by *h*-BN is close to its in-plane value because of the in-plane direction of the electric field in the *i* region. The high-frequency dielectric constant of crystalline *h*-BN in plane is $\kappa = 4.98$ and out of plane is $\kappa = 3.4$ [44]. However, the exact effective value of κ depends on the thickness of the top passivating layer. GPIN FETs could use passivating layers with very low values of κ, such as polyimide, porous BN/polyimide composites, or amorphous *h*-BN with the measured dielectric permittivity of 1.16 (close to that of air) [45–47]. In Fig. 4 we also added the plot (see the dashed curve) corresponding to the parameters similar to those of the sample d-1, but with a shorter charging time (smaller dielectric constant, $\kappa = 2.25$). As seen, the pertinent peak is shifted weaker than others. This confirms that the *i*-region geometrical capacitance markedly affects the plasmonic resonances.

The most intriguing feature of the $|P_\omega|^2$ and R_ω^V frequency dependences is the larger height of the peaks corresponding to a higher plasmonic mode. Compare the plasmonic factors at the fundamental and second resonances. The plasmonic factor $|P_\omega|^2$ near the fundamental ($\omega \simeq \Omega$) and the second ($\omega \simeq 3\Omega$) resonances (at $\eta \simeq 0$), $\nu \ll \Omega$, and $3\Omega\tau^i \ll 1$ is equal to

$$|P_\Omega|^2 \simeq \frac{(4\Omega/\pi\nu)^2}{\left|1 + \frac{8}{\pi}\left(\frac{2\sigma^i}{\sigma^g}\right)\mathcal{J}_0(\Omega t^i/2)\, e^{i\Omega t^i/2}\right|^2} \tag{25}$$

and

$$|P_{3\Omega}|^2 \simeq \frac{(12\Omega/\pi\nu)^2}{\left|1 - \frac{24}{\pi}\left(\frac{2\sigma^i}{\sigma^g}\right)\left(\frac{\Omega}{\nu}\right)^2 \mathcal{J}_0(3\Omega t^i/2) e^{i\Omega t^i/2}\right|^2}, \tag{26}$$

respectively.

Considering that the product $\sigma^g\nu$ and, therefore, the denominators on the right-hand sides of Eqs. (25) and (26) are independent of ν, one can find that $|P_\Omega|^2 \propto (\Omega/\nu)^2$ and $|P_{3\Omega}|^2 \propto (\Omega/\nu)^2$. Such dependences of the plasmonic factor peaks on ν are in line with the plots in Fig. 1(a).

As follows from Eqs. (25) and (26), the ratio of the peak heights corresponding to the fundamental and second resonances, at sufficiently large values $(2\sigma^i/\sigma^g)(\Omega/\nu)$, is equal to

$$n \simeq \left|\frac{P_{3\Omega}}{P_\Omega}\right|^2 \simeq \left[\frac{\mathcal{J}_0(\Omega t^i/2)}{\mathcal{J}_0(3\Omega t^i/2)}\right]^2. \tag{27}$$

This quantity can be smaller or larger than unity, depending on Ωt^i. For all the samples considered above (except b-3), $n > 1$; hence, the plasmonic resonances at $\omega \sim 3\Omega$ are stronger than those at $\omega \simeq \Omega$. This corresponds to the curves (a-1)–(b-2) in Fig. 2. In contrast, case b-3 is related to $n < 1$. In particular, for samples a-1 and b-1, we obtain $|P_\Omega|^2 \simeq 4.44$, $|P_{3\Omega}|^2 \simeq 42.37$, and $n \simeq 9.54$. This agrees well with the numerical calculation results shown in Fig. 2.

As mentioned above, the height of $|P_\omega|^2$ and R_ω^V resonant peaks decreases with increasing collision frequency ν. This is due to the strengthening of the plasma oscillation collisional damping. The hole and electron viscosity also damp the plasma oscillations, particularly those with larger wavenumbers q. This might be a reason for some lowering of the resonant peaks corresponding to higher plasma oscillations modes. The effect of the viscosity can be accounted for by replacing ν by $\nu^{\rm visc} = \nu + hq^2$, where h is the electron viscosity [1,32,48,49]. For the fundamental, second, and third plasma modes with $\omega_1 \simeq \Omega$, $\omega_2 \simeq 3\Omega$, and $\omega_3 \simeq 5\Omega$, the wavenumbers are $q_1 = \pi/2L$, $q_2 = 3\pi/2L$, and $q_2 = 5\pi/2L$. If $L = 1$ μm, setting $\nu = 1$–2 ps^{-1} and $h = 100$–500 cm^2/s (depending on the carrier density and the temperature [32,48,49]), $(\nu_1^{\rm visc} - \nu)/\nu \simeq 0.0123$–$0.123$ and $(\nu_2^{\rm visc} - \nu)/\nu \simeq 0.11$–$1.11$, these estimates show that the viscosity effect can decrease the heights of the second and fundamental peaks if the viscosity is sufficiently strong. As a result, the quantity $P_{3\Omega}$ (for example, for the samples a-1 and b-1) can vary from $|P_{3\Omega}|^2 \simeq 42$ at $h = 0$ to $|P_{3\Omega}|^2 \simeq 34$ and 20 at $h = 100$ cm^2/s and $h = 500$ cm^2/s, respectively. Lowering of the fundamental resonant peak is markedly smaller—about (2–12)%. However, the second resonance remains stronger even at relatively high viscosity.

In addition to the carrier collisions and the plasma viscosity mentioned above, the damping mechanisms include plasmon-plasmon interactions, radiative damping, scattering of plasmons on defects, dopant-induced plasmon decay, and the interaction of the hole and electron plasma with the carriers in highly conducting contacts (see, for example, [50–58]). The intrinsic lifetime of plasmons in GLs with the carrier densities assumed in our calculations is about 20–120 ps [51], i.e., much longer than the characteristic collision time ν^{-1}. In sufficiently short GLs, the plasmon decay time due to the contacts is of the order of ν^{-1} [54]. These estimates justify the assumption that in the GPIN FETs under consideration the collisional and viscosity damping mechanisms dominate.

As for the third resonant peaks with $\omega_3 \simeq 5\Omega$ and $q_3 \simeq 5\pi/2L$, the plasma oscillation damping due to the viscosity is strong enough to lead to the peak extinction (the right most peak in Fig. 3) at $h = 500$ cm^2/s and higher.

B. Effect of the contacts and gates coupling

As follows from Eq. (13), the plasmonic factor depends to some extent on the capacitive coupling between the side contacts and the gates. Figure 4 shows examples of the $|P_\omega|^2$ versus signal frequency $\omega/2\pi$ for the GPIN FETs with the parameters corresponding to the sample a-2 for $\eta = 0.1$ [as in Fig. 2(a)] and for $\eta = 0.2$ at $\overline{U} = 200$ mV. One can see that an increase in η leads to resonant peaks lowering. Although one needs to keep in mind that the value $\eta = 0.2$ corresponds to an overestimated contact-gate capacitance C_{cg} compared with typical values for FETs. For the comparison, we calculate $|P_\omega|^2$ also for a similar device structure, but with the short-cut side contacts and gates. In the latter case, one can use the following formula [coinciding with Eq. (20) with $\eta = 1$]:

$$|P_\omega|^2 = \frac{1}{|1 + (1/\xi_\omega)(\tan(æ_\omega L)/(æ_\omega L))|^2}. \tag{28}$$

The pertinent dependence is shown in Fig. 5 as a dotted line. As seen in the latter case, the plasma resonance associated with the gated regions (at $\omega \sim \Omega$) is suppressed, while the plasmonic resonance at $\omega \sim 2\Omega$ is sufficiently pronounced although being relatively weak.

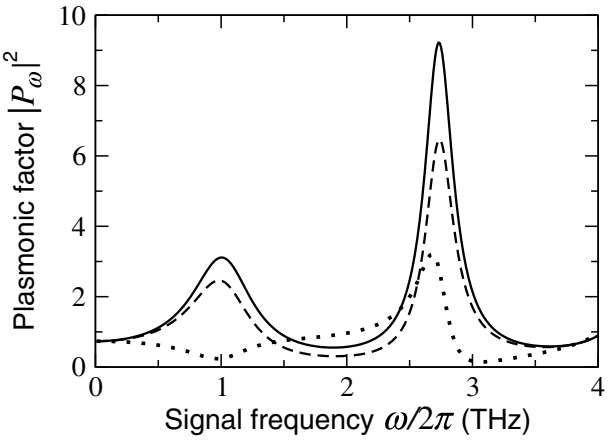

FIG. 5. Plasmonic factor $|P_\omega|^2$ versus signal frequency $f = \omega/2\pi$ for the sample with the same parameters as the sample a-2 and $\eta = 0.1$ (solid line) and $\eta = 0.2$ (dashed line). The dotted line is related to a GPIN FET with the parameters a-2, but with the short-cut side contacts and the pertinent gates.

C. Thermionic and thermogeneration currents in the reverse-biased *i* region

When $\overline{\Phi} \propto \overline{U}$ becomes relatively low, the reverse thermionic current in the reverse-biased i region can be comparable with the tunneling current. Considering the tunneling current [given by Eq. (1)] and the saturation current [given by Eq. (C2)], we find the following limitation for the minimal value of $\overline{U}$ at which the tunneling dominates over the thermionic processes:

$$e\overline{\Phi} > \left(\frac{6}{b\pi^2}\right)^{2/3}\left(\frac{2l\mu^2T^2}{\hbar v_W}\right)^{1/3}\exp\left(-\frac{2\mu}{3T}\right) = e\overline{\Phi}_{\rm min}. \tag{29}$$

For $2l = 0.2$ μm, $\mu = 100$ meV, and $T = 25$ meV, we obtain $e\overline{\Phi}_{\rm min} \simeq 17.5$ meV. This condition is satisfied in the above estimates and calculations.

The thermogeneration of the electron-hole pairs in the reverse-biased graphene p-i-n junction is primarily associated with the interband absorption of optical phonons. The thermionic rate at room temperature is estimated as $g^{\rm therm} = 10^{21}$ cm^{-2} s^{-1} [59]. For $2l = 0.2$ μm, the latter yields $J^{\rm therm} = 4elg^{\rm therm} \simeq 6.4 \times 10^{-3}$ A/cm. For comparison, the tunneling current density in the reverse-biased p-i-n junction given by Eq. (1) for the same lengths of the i region at $e\overline{\Phi} = 100$–200 meV is equal to $\overline{J}^i \simeq 0.34$–$0.95$ A/cm, i.e., one order of magnitude larger. This implies that the dark current in the GPIN FETs in the conditions under consideration is determined by the interband tunneling in the i regions.

D. Joule heating

The Joule power in the GPIN FET is equal to $Q = H\overline{J}\,\overline{V}$. For $\overline{\Phi} \lesssim \overline{V} = 100$–$200$ mV ($\overline{U} \simeq 200$–400 mV), $\overline{J} \lesssim 0.5$–$1.0$ A/cm. Assuming the device width $H = 10$ μm, for the Joule power, we obtain $Q \simeq (5$–$10)\times 10^{-5}$ W. Since the bias voltage drops primarily in the i region and somewhat around it, the Joule power releases in the area of about $2lH$. This corresponds to the thermal power density $Q/2lH$. For $2JH = 2 \times 10^{-8}$ cm^2, we obtain $Q/2lH \simeq 2.5$–5.0 kW/cm^2. The latter thermal power density is much lower than that, which can be supported by GFETs (up to 210 kW/cm^2 [60,61]). Hence, the Joule heating should not lead to a marked overheating of the GPIN-FET channel. Indeed, the heat flow from the GPIN-FET channel through, for example, the h-BN substrate can be estimated considering the h-BN thermal conductivity $k_{\rm th} \simeq 20$ W/m K [62]. Assuming the thickness of the h-BN layer $D = 1$–2 μm, for the thermal conductivity per unit area, $K_{\rm th} = k_{\rm th}/D \simeq 1$–$2$ kW/cm^{2}K. This implies that, in such a case, the Joule heating results in an increase of the channel temperature by $\Delta T \lesssim 5$ K. Since graphene has a high room-temperature thermal conductivity (about $k_{\rm th} = 5$ kW/m K) [63], the heat can also be effectively carried to the side (metallic) contacts.

E. Comparison of the Zener-Klein tunneling nonlinearity mechanisms with some other mechanisms

It appears natural to compare the mechanisms of the current rectification in the GPIN FETs under consideration and in similar detectors (for example, [9]). In the latter devices, the I-V characteristic nonlinearity in the forward-biased graphene p-n junction is used. The pertinent nonlinearity parameter $\beta^{\rm th}$ is given by Eq. (C3). Comparing the quantities β [see Eq. (17)] and $\beta^{\rm th}$ (for $\overline{\Phi} = -\overline{\Phi}^{\rm th} < 0$ in the case of the detector using the p-n nonlinearity) and $\overline{\Phi} > 0$ for the GPIN FET, we obtain

$$\frac{\beta}{\beta^{\rm th}} \simeq \frac{\pi^2 b}{8}\sqrt{\frac{\hbar v_W}{2l}}\frac{T}{\mu}\frac{\exp[(\mu - e\overline{\Phi}^{\rm th})/T]}{\sqrt{e\overline{\Phi}}}. \tag{30}$$

Setting $2l = 0.2$ μm, $\mu = 100$ meV, $\overline{\Phi} = 100$ mV, $\overline{\Phi}^{\rm th} = 25$ mV, at $T = 25$ meV ($\simeq 300$ K), we obtain $\beta/\beta^{\rm th} \simeq 0.1$. At the temperatures below room temperature, this ratio can be markedly larger. For example, at $T = (7.5$–$10)$ meV [$\simeq (90$–$120)$ K], we obtain $\beta/\beta^{\rm th} \simeq 1$–$9$.

VII. CONCLUSIONS

We demonstrate that the proposed GPIN-FETs detectors using the signal rectification due to the nonlinearity of the Zener-Klein tunneling I-V characteristics can exhibit high responsivity in the terahertz range of frequencies. The responsivity can be particularly high at the signal frequencies close to the fundamental and triple carrier plasma

frequency. This is due to the resonant excitation of plasma oscillation in the gate region of the GPIN-FET channel. The GPIN FETs can demonstrate competitive resonant responsivity at room temperatures. At lower temperatures, the responsivity can markedly increase due to the reinforcement of the resonant response because of a weakening of the carrier momentum relaxation.

ACKNOWLEDGMENTS

The Japan Society for Promotion of Science (KAKENHI Grants No. 21H04546 and No. 20K20349), Japan; RIEC Nation-Wide Collaborative Research Project No. R04/A10; the U.S. Office of Scientific Research Contract N00001435, (project monitor Dr. Ken Goretta).

The authors declare no conflict of interest related to this article.

APPENDIX A

1. Dynamic tunneling conductance of the reverse-biased *i* region

The injected holes and electrons propagating in the i region induce the current in the surrounding highly conducting (gated) regions. The dynamical conductance, apart from the dc differential conductance σ^i of the reverse-biased i region found from Eq. (1), should account for the displacement current associated with the geometrical capacitance c^i. Therefore, the dynamical conductance of the i region associated with the propagating carriers can be presented as

$$\sigma^i_\omega = \sigma^i F_\omega - i\omega c^i. \tag{A1}$$

Because of a strong nonuniformity of the electric field in the i region, the tunneling generation of the electrons and holes occurs near the edges of the p and n regions, respectively. If the generated holes and electrons propagate across the i region ballistically, their velocities are equal to $\pm v_W$. Therefore, the ac electron and hole currents are proportional to $\exp(i\omega x/v_W)$ and $\exp(-i\omega x/v_W)$. Hence, the ac current induced by the propagating carriers is proportional to

$$F_\omega = \frac{1}{2l}\int_{-l}^{l} dx g(x)[\exp(i\omega x/v_W) + \exp(-i\omega x/v_W)],$$

where, for the case of the device geometry under consideration (bladelike conducting areas), the form factor $g(x)$ is given by $g(x) = 2/\pi\sqrt{1-(x/l)^2}$ [64]. This leads to

$$F_\omega = \frac{2}{\pi} e^{i\omega t^i/2}\int_0^1 ds \frac{\cos(\omega t^i s/2)}{\sqrt{(1-s^2)}} = e^{i\omega t^i/2}\mathcal{J}_0(\omega t^i/2), \tag{A2}$$

where $t^i = 2l/v_W$ and $\mathcal{J}_0(s)$ is the Bessel function of the first kind.

If the potential drop $\overline{\Phi} > \hbar\omega_0$, the optical phonon emission can delay the holes and electrons propagation in the i region. For $\overline{\Phi}$ markedly exceeding 200 mV, the spatial dependence of the carrier velocity can be presented as $v_x = \pm v_W \exp(-x/\tau_{\rm op} v_W)$, where $\tau_{\rm op} \simeq 1$–2 ps [58] is the time of the optical phonon spontaneous emission. For the average transit time $< t^i >= (1/2l)\int_{-l}^{l} dx(2l/v_x)$, we obtain $< t^i >= 2\tau_{\rm op}\sinh(t^i/v_W\tau_{\rm op}) \simeq t^i(1 + t^i/12\tau_{\rm op}) \gtrsim t^i$. For $2l = 0.2$–$0.5\ \mu$m, the latter estimate yields $< t^i > /t^i - 1 \gtrsim (1$–$4)\%$.

The i-region capacitance c^i, which determines the capacitive component of the displacement current through the i region, for the bladelike conducting areas can be estimated, generalizing [36,65] by accounting for the presence of the highly conducting gates, as

$$c^i = \frac{\kappa}{2\pi^2}\Lambda \quad \text{or} \quad c^i = \frac{\kappa^{\rm eff}}{2\pi^2}\Lambda, \tag{A3}$$

where Λ, as in [36,52], is a logarithmic factor, which weekly depends on the geometrical parameters. Equation (A3) for the GPIN FET with $\kappa = 4.5$, yields $c^i \gtrsim 0.5$. This value is used in the main text.

APPENDIX B

In the gradual channel approximation, the spatiotemporal distributions of the electron and hole densities, $\Sigma^n(x,t)$ and $\Sigma^p(x,t)$, in the gated regions are related to the channel potential, $\varphi(x,t)$, and the gate potentials, $\varphi^n_g(t)$ and $\varphi^p_g(t)$, as

$$\Sigma^n(x,t) = \frac{\kappa}{4\pi\, ew}[\varphi^n_g(t) - \varphi(x,t)], \tag{B1}$$

$$\Sigma^p(x,t) = \frac{\kappa}{4\pi\, ew}[\varphi^p_g(t) + \varphi(x,t)], \tag{B2}$$

where κ and w are the gate layer dielectric constant and thickness. If the side source/drain contacts and the gates are coupled by a capacitive link associated with a free-space parasitic contact-gate capacitance [see Fig. 1(a)],

$$\varphi^n_g(t) = V_g + \frac{C_{cg}}{C_g + C_{cg}}\frac{U(t)}{2}, \tag{B3}$$

$$\varphi^n_g(t) = -V_g - \frac{C_{cg}}{C_g + C_{cg}}\frac{U(t)}{2}, \tag{B4}$$

where C_{cg} and C_g are the contact-gate and gate-channel capacitances, respectively.

For the dc electron and hole densities, from Eqs. (B1) and (B2) we obtain

$$\overline{\Sigma}^n = \frac{\kappa}{4\pi\, e}\left[V_g + \frac{\eta}{2}\overline{U} - \overline{\varphi}(x)\right] \simeq \frac{\kappa}{4\pi\, e}\left(V_g + \frac{\eta}{2}\overline{U}\right) = \overline{\Sigma}, \tag{B5}$$

$$\overline{\Sigma}^p = \frac{\kappa}{4\pi\, e}\left[V_g + \frac{\eta}{2}\overline{U} + \overline{\varphi}(x)\right] \simeq \frac{\kappa}{4\pi\, e}\left(V_g + \frac{\eta}{2}\overline{U}\right)\overline{\Sigma}, \tag{B6}$$

with $\eta = C_{cg}/(C_g + C_{cg})$. Since normally the dc bias voltage $\overline{V} < \overline{U} \ll V_g$, in Eqs. (B5) and (B6) we omit the spatially nonuniform terms with $\overline{\varphi}(x)$.

For the Fourier ac components, Eqs. (B1)–(B4) yield

$$\delta\Sigma^n_\omega(x) = \frac{\kappa}{4\pi\, ew}\left[\frac{\eta}{2}\delta\, V_\omega - \delta\varphi_\omega(x)\right], \tag{B7}$$

$$\delta\Sigma^p_\omega(x) = \frac{\kappa}{4\pi\, ew}\left[-\frac{\eta}{2}\delta\, V_\omega + \delta\varphi_\omega(x)\right]. \tag{B8}$$

Since the load resistance is shunt by a large capacitance and the drop of the ac potential across the load resistor is insignificant, $\delta U_\omega = \delta V_\omega$.

As usual, considering the linearized hydrodynamic equations, expressing the electron and hole average velocities, δu^n_ω and δu^p_ω, via the ac electric field $-d\delta\varphi_\omega(x)/dx$, and substituting δu^n_ω and δu^p_ω into the continuity equation, we obtain

$$\frac{d^2\delta\varphi_\omega}{dx^2} - \frac{m\omega(\omega + i\nu)}{e\overline{\Sigma}}\delta\Sigma^n = 0, \tag{B9}$$

$$\frac{d^2\delta\varphi_\omega}{dx^2} - \frac{m\omega(\omega + i\nu)}{e\overline{\Sigma}}\delta\Sigma^p = 0, \tag{B10}$$

where m is the fictitious electron/hole mass in graphene. Expressing m and $\overline{\Sigma}$ via the carrier Fermi energy, μ, in the gated region, we arrive at Eq. (9) in the main text.

If the side contacts and the gates are shortened, one can again obtain Eq. (16), but with $\eta = 1$.

Solving Eq. (9) with the boundary conditions given by Eqs. (10) and (11), we obtain

$$\delta\varphi_\omega = \frac{1}{2}\delta V_\omega\left\{\eta + (1-\eta)\cos[\text{æ}_\omega(x - l - L)] + \frac{\eta + (1-\eta)[\cos(\text{æ}_\omega L - \xi_\omega(\text{æ}_\omega L)\sin(\text{æ}_\omega L)]}{\sin(\text{æ}_\omega L) + \xi_\omega(\text{æ}_\omega L)\cos(\text{æ}_\omega L)]} \times \sin[\text{æ}_\omega(x - l - L)]\right\}, \tag{B11}$$

$$\delta\varphi_\omega = -\frac{1}{2}\delta V_\omega\left\{\eta + (1-\eta)\cos[\text{æ}_\omega(x + l + L)] - \frac{\eta + (1-\eta)[\cos(\text{æ}_\omega L) - \xi_\omega(\text{æ}_\omega L)\sin(\text{æ}_\omega L)]}{\sin(\text{æ}_\omega L) + \xi_\omega(\text{æ}_\omega L)\cos(\text{æ}_\omega L)} \times \sin[\text{æ}_\omega(x + l + L)]\right\} \tag{B12}$$

for the n region and p region, respectively. Here $\text{æ}_\omega = \pi\sqrt{\omega(\omega + i\nu)}/2\Omega L$, $\xi_\omega = \sigma^g_\omega/2\sigma^i_\omega = (\sigma^g/2\sigma^i_\omega)[i\nu/(\omega + i\nu)] = (\sigma^g/2\sigma^i)[i\nu/(\omega + i\nu)(F_\omega - i\omega\tau^i)]$, and

$$\Omega = \frac{\pi\, e}{\hbar L}\sqrt{\frac{\mu\, w}{\kappa}} \tag{B13}$$

is the plasma frequency of the gated regions.

APPENDIX C

The thermionic dc current density, including both the hole and electron components, in the i region of the GPIN FETs under consideration can be presented as

$$\overline{J}^{\text{th}} = J_s \exp\left(-\frac{\mu}{T}\right)\left[\exp\left(-\frac{e\overline{\Phi}^{\text{th}}}{T}\right) - 1\right], \tag{C1}$$

where

$$J_s \simeq \frac{4eT\mu}{\pi^2\hbar^2 v_W}\exp\left(-\frac{\mu}{T}\right) \tag{C2}$$

is the graphene p-i-n junction saturation current density. In Eq. (C1), as in the main text, the p-i-n junction reverse and forward bias voltages $\overline{\Phi} > 0$ and $\overline{\Phi}^{\text{th}} < 0$, respectively.

Equations (C1) and (C2) yield the following value of the I-V characteristic nonlinearity parameter:

$$\beta^{\text{therm}} = \frac{1}{2}\frac{d^2\overline{J}^{\text{therm}}}{d\overline{\Phi}^2} = \frac{\sigma^{i,\text{th}}}{2T} \simeq \frac{2e^3}{\pi^2\hbar^2 v_W}\left(\frac{\mu}{T}\right)\exp\left(-\frac{\mu + e\overline{\Phi}}{T}\right). \tag{C3}$$

Here $\sigma^{i,\text{th}}$ is the differential conductance.

[1] M. I. Dyakonov and M. S. Shur, Plasma wave electronics: novel terahertz devices using two-dimensional electron fluid, IEEE Trans. Electron. Devices **43**, 1640 (1996).

[2] V. V. Popov, N. Pala, and M. S. Shur, Room temperature terahertz plasmonic detection by antenna arrays of field-effect transistors, Nanosci. Nanotechnol. Lett. **4**, 1015 (2012).

[3] T. Otsuji, T. Watanabe, S. A. Boubanga Tombet, A. Satou, W. M. Knap, V. V. Popov, M. Ryzhii, and V. Ryzhii, Emission and detection of terahertz radiation using

two-dimensional electrons in III–V semiconductors and graphene, IEEE Trans. Terahertz Sci. Technol. **3**, 63 (2013).
[4] T. Otsuji, T. Watanabe, S. A. Boubanga Tombet, A. Satou, V. Ryzhii, V. Popov, and W. Knap, Emission and detection of terahertz radiation using two-dimensional plasmons in semiconductor nanoheterostructures for nondestructive evaluations, Opt. Eng. **53**, 031206 (2014).
[5] L. Vicarelli, M. S. Vitiello, D. Coquillat, A. Lombardo, A. C. Ferrari, W. Knap, M. Polini, V. Pellegrini, and A. Tredicucci, Graphene field-effect transistors as room temperature terahertz detectors, Nat. Mat. **11**, 865 (2012).
[6] A. Zak, M. A. Andersson, M. Bauer, J. Matukas, A. Lisauskas, H. G. Roskos, and J. Stake, Antenna-integrated 0.6 THz FET direct detectors based on CVD graphene, Nano Lett. **14**, 5834 (2014).
[7] V. Ryzhii, M. Ryzhii, M. S. Shur, V. Mitin, A. Satou, and T. Otsuji, Resonant plasmonic terahertz detection in graphene split-gate field-effect transistors with lateral p-n junctions, J. Phys. D: Appl. Phys. **49**, 315103 (2016).
[8] D. Yadav, S. Boubanga-Tombet, T. Watanabe, S. Arnold, V. Ryzhii, and T. Otsuji, Terahertz wave generation and detection in double-graphene layered van der Waals heterostructures, 2D Mater. **3**, 045009 (2016).
[9] D. V. Fateev, K. V. Mashinsky, and V. V. Popov, Terahertz plasmonic rectification in a spatially periodic graphene, Appl. Phys. Lett. **110**, 0611106 (2017).
[10] I. A. Gayduchenko, G. E. Fedorov, M. V. Moskotin, D. I. Yagodkin, S. V. Seliverstov, G. N. Goltsman, A. Yu. Kuntsevich, M. G. Rybin, E. D. Obraztsova, V. G. Leiman *et al.*, Manifestation of plasmonic response in the detection of sub-terahertz radiation by graphene based devices, Nanotechnology **29**, 245204 (2018).
[11] D. A. Bandurin, D. Svintsov, I. Gayduchenko, S. G. Xu, A. Principi, M. Moskotin, I. Tretyakov, D. Yagodkin, S. Zhukov, T. Taniguchi, K. Watanabe, I. V. Grigorieva, M. Polini, G. N. Goltsman, A. K. Geim, and G. Fedorov, Resonant terahertz detection using graphene plasmons, Nat. Commun. **9**, 5392 (2018).
[12] V. Ryzhii, T. Otsuji, and M. S. Shur, Graphene based plasma-wave devices for terahertz applications, Appl. Phys. Lett. **116**, 140501 (2019).
[13] P. R. Wallace, Band theory of graphite, Phys. Rev. **71**, 622 (1947).
[14] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, The electronic properties of graphene, Rev. Mod. Phys. **81**, 109 (2009).
[15] T. Low and Ph. Avouris, Graphene plasmonics for terahertz to mid-infrared applications, ACS Nano, **8**, 1086 (2014).
[16] Y. Li, K. Tantiwanichapan, A. K. Swan, and R. Paiella, Graphene plasmonic devices for terahertz optoelectronics, Nanophotonics **9**, 1901 (2020).
[17] V. V. Cheianov and V. I. Fal'ko, Selective transmission of Dirac electrons and ballistic magnetoresistance of *n-p* junctions in graphene, Phys. Rev **74**, 041103 (2006).
[18] A. Ossipov, M. Titov, and C. W. J. Beenakker, Reentrance effect in a graphene *n-p-n* junction coupled to a superconductor, Phys. Rev. B **75**, 241401(R) (2007).
[19] D. Jena, Tunneling transistors based on graphene and 2-D crystals, Proc. IEEE **101**, 1585 (2013).
[20] N. Vandecasteele, A. Barreiro, M. Lazzeri, A. Bachtold, and F. Mauri, Current-voltage characteristics of graphene devices: Interplay between Zener-Klein tunneling and defects, Phys. Rev. B **82**, 045416 (2010).
[21] V. Ryzhii, M. Ryzhii, V. Mitin, and M. S. Shur, Graphene tunneling transit-time terahertz oscillator based on electrically induced p-i-n junction, Appl. Phys. Exp. **2**, 034503 (2009).
[22] V. L. Semenenko, V. G. Leiman, A. V. Arsenin, V. Mitin, M. Ryzhii, T. Otsuji, and V. Ryzhii, Effect of self-consistent electric field on characteristics of graphene p-i-n tunneling transit-time diodes, J. Appl. Phys. **113**, 024503 (2013).
[23] J. C. Song, D. A. Abanin, and L. S. Levitov, Coulomb drag mechanisms in graphene, Nano Lett. **13**, 3631 (2013).
[24] M. Schütt, P. M. Ostrovsky, M. Titov, I. V. Gornyi, B. N. Narozhny, and A. D. Mirlin, Coulomb Drag in Graphene Near the Dirac Point, Phys. Rev. Lett. **110**, 026601 (2013).
[25] R. V. Gorbachev, A. K. Geim, M. I. Katsnelson, K. S. Novoselov, T. Tudorovskiy, I. V. Grigorieva, A. H. MacDonald, S. V. Morozov, K. Watanabe, T. Taniguchi, and L. A. Ponomarenko, Strong Coulomb drag and broken symmetry in double-layer graphene, Nat. Phys. **8**, 896 (2012).
[26] V. Ryzhii, M. Ryzhii, V. Mitin, M. S. Shur, and T. Otsuji, S-Shaped Current-Voltage Characteristics of n^+-*i*-*n*-n^+ Graphene Field-Effect Transistors Due the Coulomb Drag of Quasi-Equilibrium Electrons by Ballistic Electrons, Phys. Rev. Appl. **16**, 014001 (2021).
[27] V. Ryzhii, M. Ryzhii, A. Satou, T. Otsuji, V. Mitin, and M. S. Shur, Effect of Coulomb Carrier Drag and Terahertz Plasma Instability in p^+-*p*-*i*-*n*-n^+ Graphene Tunneling Transistor Structures, Phys. Rev. Appl. **16**, 064054 (2021).
[28] V. Ryzhii, M. Ryzhii, V. Mitin, M. S. Shur, and T. Otsuji, Coulomb electron drag mechanism of terahertz plasma instability in n^+-i-n-n^+ graphene FETs with ballistic injection, Appl. Phys. Lett. **119**, 093501 (2019).
[29] M. S. Shur and L. F. Eastman, Ballistic transport in semiconductor at low temperatures for low-power high-speed logic, IEEE Trans. Electron Devices **26**, 1677 (1979).
[30] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, Theory of ballistic nanotransistors, IEEE Trans. Electron Devices **50**, 1853 (2003).
[31] G. Liang, N. Neophytou, D. E. Nikonov, and M. S. Lundstrom, Performance projections for ballistic graphene nanoribbon field-effect transistors, IEEE Trans. Electron Devices **54**, 677 (2007).
[32] Y. Zhang and M. S. Shur, Collision dominated, ballistic, and viscous regimes of terahertz plasmonic detection by graphene, J. Appl. Phys. **129**, 053102 (2021).
[33] X. Li, E. A. Barry, J. M. Zavada, M. Buongiorno Nardelli, and K. W. Kim, Influence of electron-electron scattering on transport characteristics in monolayer graphene, Appl. Phys. Lett. **97**, 082101 (2010).
[34] A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanabe, T. Taniguchi, and A. K. Geim, Micrometer-scale

ballistic transport in encapsulated graphene at room temperature, Nano Lett. **11**, 2396 (2011).
[35] L. Banszerus, M. Schmitz, S. Engels, M. Goldsche, K. Watanabe, T. Taniguchi, B. Beschoten, and C. Stampfer, Ballistic transport exceeding 28 μm in CVD grown graphene, Nano Lett. **16**, 1387 (2016).
[36] S. G. Petrosyan and A. Ya. Shik, Contact phenomena in low-dimensional electron systems, Sov. Phys. -JETP **69**, 2119 (1989).
[37] B. L. Gelmont, M. S. Shur, and C. Moglestue, Theory of junction between two-dimensional electron gas and p-type semiconductor, IEEE Trans. Electron Devices **39**, 1216 (1992).
[38] D. B. Chklovskii, B. I. Chklovskii, and L. I. Glasman, Electrostatics of edge channels, Phys. Rev. B **46**, 4026 (1992).
[39] V. Ryzhii, A. Satou, I. Khmyrova, M. Ryzhii, T. Otsuji, V. Mitin, and M. S. Shur, Plasma effects in lateral Schottky junction tunneling transit-time terahertz oscillator, J. Phys. Conf. Ser. **38**, 228 (2006).
[40] V. Ryzhii, Terahertz plasma waves in gated graphene heterostructures, Jpn. J. Appl. Phys. **45**, L923 (2006).
[41] L. A. Falkovsky and A. A. Varlamov, Space-time dispersion of graphene conductivity, European Phys. J. B **56**, 281 (2007).
[42] V. Ryzhii, A. Satou, and T. Otsuji, Plasma waves in two-dimensional electron-hole system in gated graphene, J. Appl. Phys. **101**, 024509 (2007).
[43] A. Soltani, F. Kuschewski, M. Bonmann, A. Generalov, A. Vorobiev, F. Ludwig, Matthias M. Wiecha, D. Čibiraitė, F. Walla, S. Winnerl *et al.*, Direct nanoscopic observation of plasma waves in the channel of a graphene field-effect transistor, Light Sci. Appl. **9**, 1 (2020).
[44] A. Pierret, D. Mele, H. Graef, J. Palomo, T. Taniguchi, K. Watanabe, Y. Li, B. Toury, C. Journet, and P. Steyer *et al.*, Dielectric permittivity, conductivity and breakdown field of hexagonal boron nitride, Mater. Res. Express **9**, 065901 (2022).
[45] P. Zhang, L. Zhang, Ke Zhang, J. Zhao, and , Yao Li, Preparation of polyimide films with ultra-low dielectric constant by phase inversion, Crystals **11**, 1383 (2021).
[46] K. Yang, Yi Young Kang, H. Jeong Ahn, D.-F. Kim, No Kyun Park, S. Q. Choi, J. Chan Won, and Y. Ho Kim, Porous boron nitride/polyimide composite films with high thermal diffusivity and low dielectric properties via high internal phase Pickering emulsion method, Ind.Eng. Chem. **82**, 173 (2020).
[47] S. Hong, C.-S. Lee, M.-H. Lee, Y. Lee, K. Y. Ma, G. Kim, S. I. Yoon, K. Ihm, K.-J. Kim, T. J. Shin *et al.*, Ultralow-dielectric-constant amorphous boron nitride, Nature **582**, 511 (2020).
[48] A. I. Berdyugin, S. G. Xu, F. M. D. Pellegrino, R. K. Kumar, A. Principi, I. Torre, M. B. Shalom, T. Taniguchi, K. Watanabe, I. V. Grigorieva, M. Polini, A. K. Geim, and D. A. Bandurin, Hall viscosity of graphene's electron fluid, Science **364**, 162 (2019).
[49] A. Principi, G. Vignale, M. Carrega, and M. Polini, Bulk and shear viscosities of the two-dimensional electron liquid in a doped graphene sheet, Phys. Rev. B **93**, 125410 (2016).
[50] H. Yan, T. Low, W. Zhu, Y. Wu, M. Freitag, X. Li, F. Guinea, P. Avouris, and F. Xia, Damping pathways of mid-infrared plasmons in graphene nanostructures, Nat. Photonics **7**, 394 (2013).
[51] A. Principi, G. Vignale, M. Carrega, and M. Polini, Intrinsic lifetime of Dirac plasmons in graphene, Phys. Rev. B **88**, 195405 (2013).
[52] V. Semenenko, S. Schuler, A. Centeno, A. Zurutuza, T. Mueller, and V. Perebeinos, Plasmon–plasmon interactions and radiative damping of graphene plasmons, ACS Photonics **5**, 3459 (2018).
[53] J. L. Garcia-Pomar, A. Yu. Nikitin, and L. Martin-Moreno, Scattering of graphene plasmons by defects in the graphene sheet, ACS Nano **7**, 4988 (2013).
[54] A. Satou, V. Ryzhii, F. T. Vasko, V. V. Mitin, and T. Otsuji, Numerical simulation of terahertz plasmons in gated graphene structures, Proc. SPIE, **8624**, 37 (2013).
[55] A. Satou, Y. Koseki, V. Vyurkov, and T. Otsuji, Damping mechanism of terahertz plasmons in graphene on heavily doped substrate, J. Appl. Phys. **115**, 104501 (2014).
[56] M. R. Singh, M. J. Brzozowski, and B. Apter, Effect of phonon-plasmon and surface plasmon polaritons on photoluminescence in quantum emitter and graphene deposited on polar crystals, J. Appl. Phys. **120**, 124308 (2016).
[57] M. J. Brzozowski and M. R. Singh, Photoluminescence quenching in quantum emitter, metallic nanoparticle, and graphene hybrids, Plasmonics **12**, 1021 (2017).
[58] D. Novko, Dopant-induced plasmon decay in graphene, Nano Lett. **17**, 6991 (2017).
[59] F. Rana, P. A. George, J. H. Strait, S. Shivaraman, M. Chandrashekhar, and M. G. Spencer, Carrier recombination and generation rates for intravalley and intervalley phonon scattering in graphene, Phys. Rev. B **79**, 115447 (2009).
[60] M. Freitag, H.-Y. Chiu, M. Steiner, V. Perebeinos, and P. Avouris, Thermal infrared emission from biased graphene, Nat. Nanotechnol. **5**, 497 (2008).
[61] M. Freitag, M. Steiner, Y. Martin, V. Perebeinos, Z. Chen, J. C. Tsang, and P. Avouris, Energy dissipation in graphene field-effect transistors, Nano Lett. **9**, 1883 (2009).
[62] J.-C. Zheng, L. Zhang, A. V. Kretinin, S. V. Morozov, Yi BoWang, T. Wang, X. Li, F. Ren, J. Zhang, C.-Yu Lu, J.-C. Chen, M. Lu, H.-Q. Wang, A. K. Geim, and K. Novoselov, High thermal conductivity of hexagonal boron nitride laminates, 2D Mater. **3**, 011004 (2016).
[63] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, and C. Ning Lau, Superior thermal conductivity of single-layer graphene, Nano Lett. **8**, 902 (2008).
[64] V. Ryzhii and G. Khrenov, High-frequency operation of lateral hot-electron transistor, IEEE Trans. Electron Devices **42**, 166 (1995).
[65] A. Sh. Achoyan, A. É. Yesayan, É. M. Kazaryan, and S. G. Petrosyan, Two-dimensional *p-n* junction under equilibrium conditions, Semiconductors **36**, 903 (2002).