Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Interchannel Coupling in Two-Channel THz Detectors]{Interchannel Coupling in Two-Channel AlGaN/GaN Plasmonic Terahertz Detectors

Srinivasa Ramanujam
SASTRA University
Sabareesan P
SASTRA University
Sathish Kumar Dhayalan
Indian Space Research Organisation
Muthubalan Varadharajaperumal
muthubalan@ece.sastra.edu

SASTRA University

Research Article

Keywords:

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11304022/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Interchannel Coupling in Two-Channel AlGaN/GaN Plasmonic Terahertz Detectors

Srinivasa Ramanujam[1], Sabareesan. P[1], Sathish Kumar Dhayalan[2], Muthubalan Varadharajaperumal[1*]

[1*]SASTRA Deemed University, Thanjavur, 613401, Tamil Nadu, India.
[2]Semi-Conductor Laboratory, Sector 72, S.A.S. Nagar, Mohali, 160071, Punjab, India.

*Corresponding author(s). E-mail(s): muthubalan@ece.sastra.edu;

Abstract

A second two-dimensional electron gas (2DEG) forms in AlGaN/GaN high-electron-mobility transistors (HEMTs) capped with AlN. The strong response of measured double-channel terahertz (THz) detectors has been attributed to the coupling between the channels, but a measured responsivity also contains the antenna coupling and the bias, so whether the coupling itself helps is not known. We isolate the coupling by modelling plasma-wave rectification in this detector and in uncapped and GaN-capped single-channel references. A semi-analytical theory, to our knowledge the first for rectification in asymmetric coupled channels, joins analytic coupled eigenmodes of the gated region to transmission-line access regions and splits the photovoltage exactly into modal terms. A full-device perturbation model adds Schrödinger–Poisson and two-dimensional electrostatics, a floating drain and hot electrons. A nonlinear time-domain solver verifies its implementation within 3.0% over 375 cases, the model reproduces the Dyakonov–Shur limit within 0.4%, and a Boltzmann (Mermin) estimate keeps kinetic corrections below 0.3% at zero bias. We find that in these AlN-capped devices the coupling does not help. For matched, conjugate or direct antenna coupling and a floating drain, the single-channel devices respond best (0.22–0.28 V/W at 1 THz, against 0.10 V/W). A design map over upper-channel mobility and barrier thickness shows that a better upper channel makes the detector worse: it shunts the access regions and lowers the device impedance, while the rectification itself changes little. Toward threshold the small-signal noise-equivalent power falls to about 20–23 pW/$\sqrt{\text{Hz}}$ in the hydrodynamic model. Both models apply to other stacks of capacitively coupled channels.

Keywords: THz detection, Plasma-wave rectification, AlGaN/GaN HEMT, coupled plasmons, hot electrons, device simulation

1 Introduction

Plasma-wave rectification in field-effect transistors, proposed by Dyakonov and Shur [1, 2] and observed resonantly by Knap et al. [3], underlies THz detectors in CMOS, III–V and graphene technologies [4], most of them with a single two-dimensional electron gas (2DEG). AlGaN/GaN HEMTs, with high densities and long relaxation times [5], often have a second one: a strongly polarizing AlN cap induces a 2DEG above the interface channel [6], and double-channel HEMTs are built on purpose for RF/power [7] and THz detection [8]. Meng et al. credited the high responsivity of their double-channel detector to the interaction of the two 2DEGs [8]. The question

matters for detector design: if the coupling helps, a second channel is a design tool; if not, it is a parasitic to avoid. The second channel adds an acoustic-like antiphase plasmon and rectification terms that mix the two modes, but whether this helps detection has not been settled, and a measurement cannot easily tell, because a measured responsivity also contains the antenna coupling and the gate bias.

The linear plasmon physics of coupled 2DEGs is well studied: absorption in GaN double-channel HEMTs [9], instability-driven acoustic and optical modes [10], gated bilayers [11] and double-quantum-well transistors [12, 13]. Their rectification has received less attention. Double-graphene-layer detectors [14] rely on symmetric coupling and on tunneling or interband nonlinearities, and the TCAD study of a double-channel detector that we are aware of is non-resonant [15]. Full-wave hydrodynamic simulations of multichannel grating-gate GaN HEMTs predict deeper absorption resonances as channels are added [16], but absorption is not the rectified photovoltage. Single channels are better understood: non-resonant detection and readout loading [17, 18], distributed self-mixing [19], viscous [20] and thermoelectric [21] contributions, and geometry outside the gated channel [22].

We know of no semi-analytical theory of rectification in asymmetric, capacitively coupled channels, and no model that isolates the coupling in a complete device. This paper provides both and uses them to answer the question. (i) A coupled-eigenmode theory, with densities from a self-consistent electrostatic solution, joins the gated region to transmission-line access regions and splits the photovoltage exactly into modal terms (Section 2). (ii) A full-device perturbation model with 2D electrostatics, a floating drain and hot electrons is verified against a nonlinear time-domain solver and the Dyakonov–Shur limit (Sections 3–4). (iii) In AlN-capped HEMTs at zero bias the second channel always lowers the response, because it shunts the access regions, and a better upper channel makes it worse; depleting it recovers only part of the loss (Section 5).

2 Device and semi-analytical theory

The three AlGaN/GaN HEMTs are identical up to the top of a 25 nm $Al_{0.25}Ga_{0.75}N$ barrier (Fig. 1). No-cap has nothing above it, GaN-cap has a 5 nm doped GaN cap and AlN-cap a 5 nm doped AlN cap. In AlN-cap the positive AlN/AlGaN polarization charge accumulates a second 2DEG (channel a) on the AlGaN side [6, 23], coupled to the interface channel b. All three have $L_g = 0.3\,\mu$m, $L_{sg} = 1.0\,\mu$m, $L_{gd} = 2.0\,\mu$m, $W = 20\,\mu$m and a 1.5 μm buffer on SiC (Table 1).

2.1 Electrostatics and statistics

We obtain the equilibrium densities from a self-consistent Schrödinger–Poisson solution along the growth direction, with a common Fermi level at zero bias. The charges are the polarization sheets (Table 1; relative to GaN [23, 24]), the ionized cap donors, deep donor-like traps in the barrier and buffer (the buffer traps half compensated by acceptors), and the electrons. The electron density comes from the effective-mass Schrödinger equation with the band masses and Kane nonparabolicity α of Table 1, weighted as in Eq. (1) by $|\psi_i|^2$; a damped predictor–corrector iteration [25] converges the loop (Fig. 2a). Under the gate the Schottky barrier is $\phi_B = (1.3x + 0.84)\,$eV for a top-layer Al fraction x [26]. In the access regions the free surface carries donors of density N_{TS} at depth E_{TS} below the conduction band, and their ionized charge neutralizes the stack [24]. With no back plane, the buffer has zero field at its base.

In the transport equations each 2DEG is described by its chemical potential $\eta(n) = E_F - E_c$ at the interface, taken from multi-subband Fermi–Dirac statistics with the same masses and non-parabolicity,

$$n = \sum_i \mathcal{D}\Big[k_B T \ln\big(1 + e^{x_i}\big) + 2\alpha (k_B T)^2 F_1(x_i)\Big],$$
$$x_i = \frac{\eta - E_i}{k_B T},$$
$$E_i = c_E E_0^{\mathrm{Ai}}(n) + s_S\big[E_i^{\mathrm{Ai}}(n_{\mathrm{nom}}) - E_0^{\mathrm{Ai}}(n_{\mathrm{nom}})\big], \tag{1}$$

where $\mathcal{D} = m_b/(\pi\hbar^2)$, F_1 is the Fermi–Dirac integral and $E_i^{\mathrm{Ai}}(n) = (\hbar^2/2m_b)^{1/3}[3\pi q^2 n(i + \frac{3}{4})/(4\varepsilon)]^{2/3}$ are the triangular-well (Airy) levels [26]. The subband spacing ($s_S = 0.66$ and 0.53)

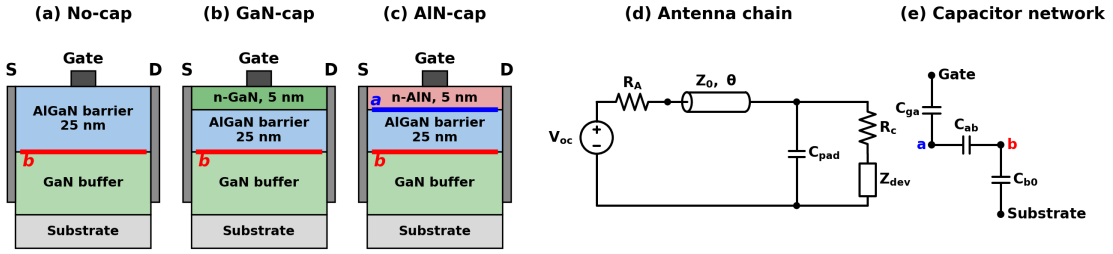


Fig. 1 (a)–(c) Device cross-sections (not to scale). Channel b (red) forms at the AlGaN/GaN interface in all devices; channel a (blue) forms only in AlN-cap (channel colours, not device colours); n-GaN and n-AlN are the doped caps. (d) Antenna chain: Thévenin antenna ($V_{oc} = E_0 l_{\text{eff}}$, R_A), matching line (Z_0, electrical length θ), antenna-feed pad capacitance C_{pad}, source contact resistance R_c and device gate–source impedance Z_{dev}. (e) Capacitor network of the semi-analytical theory.

Table 1 Device parameters. Derived equilibrium quantities (zero bias, 300 K) are from Sections 2.1–2.2.

Parameter	No-cap	GaN-cap	AlN-cap (channel a / b)
L_g, L_{sg}, L_{gd}, W	300 nm, 1.0 μm, 2.0 μm, 20 μm		
Barrier t_{AlGaN}, x; buffer; gate metal	25 nm, 0.25; 1.5 μm GaN on SiC; 200 nm		
Cap thickness, doping	—	5 nm GaN, 5×10^{18} cm^{-3}	5 nm AlN, 5×10^{18} cm^{-3}
Polarization sheets (cm^{-2})	AlGaN/GaN 1.385×10^{13}; AlN/AlGaN 4.84×10^{13}		
Schottky barrier ϕ_B (eV)	1.165	0.84	2.14
Surface donors N_{TS}, E_{TS}	5×10^{13} cm^{-2}, 1.3 eV		
Buffer traps N_{TB}, E_{TB}	5×10^{17} cm^{-3} (half compensated), 0.58 eV		
Barrier traps N_{TBA}, E_{TBA}	5×10^{16} cm^{-3}, 0.33 eV		
Band mass (GaN, AlGaN, AlN)	0.20, 0.23, 0.32 m_0		
Non-parabolicity α (GaN, AlGaN)	0.187, 0.152 eV^{-1}		
$\varepsilon_{\text{GaN,AlN,sub}}$, T	9.5, 8.5, 9.7; 300 K		
τ_E, R_c, v_{sat}, E_0	0.5 ps; 0.3 Ωmm; 1.9×10^5 m/s; 10^4 V/m		
Gated, access density (10^{13} cm^{-2})	1.02, 1.00	0.86, 0.79	3.60, 3.87 / 1.25, 1.25
E_F above ground subband (k_BT); m^*/m_b	4.3; 1.05	3.7; 1.04	12.0; 1.08 / 5.1; 1.05
μ (cm^2/Vs)	1500	1500	45 / 1500
γ (s^{-1})	5.6×10^{12}	5.6×10^{12}	1.57×10^{14} / 5.6×10^{12}
τ_{ee}, τ_2 (fs)	45, 36	40, 32	105, 6.0 / 52, 40
ν (m^2/s)	1.9×10^{-3}	1.5×10^{-3}	0.6×10^{-3} / 2.3×10^{-3}
R_{out} (Ω)	100	121	99

comes from the Schrödinger–Poisson solution. The centroid coefficient c_E (0.012 and 0.59 for the GaN and AlGaN wells) is fitted so that $\eta(n)$ reproduces the Schrödinger–Poisson gate capacitance $q\,dn/dV_G$; applied unchanged to GaN-cap, the fit gives its capacitance within 0.4%. One function thus covers degeneracy, the charge centroid and the non-degenerate limit, and in the dynamics $\eta' = d\eta/dn$ acts as a quantum-plus-centroid capacitance q^2/η'.

Table 1 lists the zero-bias gated and access densities (AlN-cap channels split at the density minimum in the barrier; Fig. 2a,b). An unfitted triangular-well model ($c_E = 1$, parabolic band) gives gated densities of 0.98, 0.83 and $3.28/1.39 \times 10^{13}$ cm^{-2}, so quantization raises channel a's density by about a tenth and lowers channel b's by as much. The GaN cap *lowers* the density. Its negative polarization charge at the cap/barrier interface raises the conduction band at the top of the barrier to 2.0 eV above E_F, and the cap donors (2.5×10^{12} cm^{-2}) do not make up for it. The gate fixes the surface potential, so surface donors affect only the access density; in AlN-cap they are fully ionized and cap the total access density. The first excited subband lies 0.14 eV (GaN wells) and 0.27 eV (channel a) above the ground state and holds 4–10% of the interface electrons and 13% of channel a's. Intersubband scattering is fast, so we treat each channel as one fluid. Gate bias depletes the single-channel devices near −5.6 V; in AlN-cap channel a depletes near −4.8 V and channel b near −13 V (Fig. 2c).

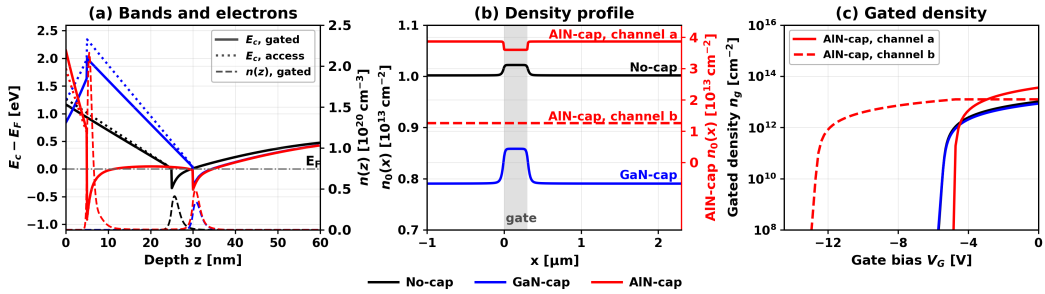


Fig. 2 Device colours throughout: No-cap black, GaN-cap blue, AlN-cap red. (a) Schrödinger–Poisson conduction band under the gate (solid) and in the access region (dotted) at zero bias, with the electron density under the gate (dashed, right axis). (b) Equilibrium density along the channel (gate shaded); No-cap and GaN-cap on the left axis, AlN-cap (channel a solid, channel b dashed) on the right axis. (c) Gated density vs. gate bias.

2.2 Transport equations

We treat each channel as a one-dimensional electron fluid with sheet density n, velocity v and electron-temperature rise $\Theta = T_e - T$:

$$\partial_t n + \partial_x(nv) = 0, \tag{2a}$$

$$\partial_t v + v\partial_x v = \frac{q}{m^*}\partial_x u - \frac{\partial_x(b\Theta)}{m^* n} - \gamma(n,\Theta)v + \frac{\partial_x(n\nu\partial_x v)}{n}, \tag{2b}$$

$$C_v(\partial_t\Theta + v\partial_x\Theta) + \beta\partial_x v = m^* n\left[\gamma_0 v^2 + \nu(\partial_x v)^2\right] - \frac{C_v\Theta}{\tau_E} + \partial_x(\kappa\partial_x\Theta), \tag{2c}$$

Here $u = \phi - [\eta(n) - \eta(n_0)]/q$ is the electrochemical force potential (ϕ: electrostatic potential) and m^* the optical (plasma) mass of the non-parabolic band, $m^* = m_b(1 + 2\alpha\bar{\varepsilon})$ with $\bar{\varepsilon} = 2P/n$ ($= E_F - E_i$ when degenerate). (Table 1 gives m^*/m_b.) Equation (2c) is the electron heat equation [27] with coefficients frozen at equilibrium. The heat capacity $C_v = \partial E/\partial T|_n$, compression coefficient $\beta = T\,\partial P/\partial T|_n$ and thermal-force coefficient $b = \partial P/\partial T|_n$ come from the kinetic pressure P of Eq. (1). The remaining terms are heating by friction ($\gamma_0 = \gamma(n_0, T)$) and viscosity, energy relaxation with $\tau_E = 0.5$ ps [28], and Wiedemann–Franz conduction, $\kappa = (\pi^2/3)(k_B/q)^2 Tqn\mu$. This Lorenz number is an upper bound, since e–e collisions relax the heat current but not the charge current (Section 2.3). For a single degenerate subband $\beta/C_v = T$. The populated excited subbands add heat capacity and lower this to 0.36–0.47 T in the interface channels and 0.55 T in channel a. At THz frequencies ($\omega\tau_E = 3$) the electron temperature therefore follows the density adiabatically, $\Theta \approx (0.36\text{–}0.55)\,T\,\delta n/n_0$.

Mobility. Scattering mechanisms are combined with Matthiessen's rule. In the interface 2DEGs they are polar-optical phonon scattering, $\mu_{\mathrm{POP}} \propto e^{\hbar\omega_{LO}/k_BT} - 1$ ($\hbar\omega_{LO} = 92$ meV) [29]; a temperature-independent background (7000 cm^2/Vs) chosen to give 1500 cm^2/Vs at 300 K; and remote scattering from the doped cap, $\mu_R = 8q(k_F d)^3/(\pi\hbar n_R)$ [30], with d the cap-to-channel distance and n_R the ionized cap-donor sheet density. Channel a lies in the $Al_{0.25}Ga_{0.75}N$ alloy, where alloy disorder limits the room-temperature mobility of 2DEGs at these densities to below 50 cm^2/Vs [6]. Alloy scattering in the Fang–Howard well [31] (alloy potential 1.8 eV, $\mu \propto n^{-1/3}$) gives only about 310 cm^2/Vs, so we add a constant term, standing for the alloy disorder this simple estimate misses, that brings the total to 45 cm^2/Vs at nominal density. The rate $\gamma = q/(m^*\mu)$ plus a capture-rate estimate of trap scattering ($\lesssim 1.5\%$ under the gate, up to 32% in the access regions) depends on density, $\gamma_n = \partial\gamma/\partial n$, and on electron temperature. For the hot-electron coefficient $\gamma_T = \partial\gamma/\partial T_e$ (lattice at T) we use the energy-resolved polar-optical-phonon rate with Pauli blocking, $1/\tau(E) \propto N_q[1 - f(E + \hbar\omega_{LO})]/[1 - f(E)] + (N_q + 1)H(E - \hbar\omega_{LO})[1 - f(E - \hbar\omega_{LO})]/[1 - f(E)]$ (H: step function; energy-independent 2D matrix element, f Fermi–Dirac at T_e, N_q Bose at T), averaged as $\gamma = \int E(-\partial_E f)dE / \int E(-\partial_E f)\tau\,dE$ and scaled to γ_{POP}. At nominal density $\gamma_T/\gamma_{\mathrm{POP}} = 3.9\text{–}5.4 \times 10^{-3}\,\mathrm{K}^{-1}$, 8–34% below the Fermi-surface limit $\hbar\omega_{LO}/(2k_BT^2)$, and the ratio falls further toward threshold.

Viscosity. We use $\nu = \tau_2 P/(m^* n)$ ($= v_F^2\tau_2/4$ when degenerate) with $1/\tau_2 = 1/\tau_{ee} + \gamma$ [32], where the 2D Fermi-liquid electron–electron time [33, 34]

is

$$\frac{\hbar}{\tau_{ee}} = \frac{\pi}{4}\frac{(k_B T)^2}{E_F}\Big[\ln\frac{E_F}{k_B T} + \ln\frac{2q_{TF}}{k_F} + 1\Big] \quad (3)$$

(q_{TF}: Thomas–Fermi wavenumber), multiplied by $1 - e^{-n/\mathcal{D}k_B T}$ and with $E_F \to \max(E_F, k_B T)$ to cover the non-degenerate limit. With τ_{ee} and ν of Table 1, τ_2 is 4.5–5.5 times shorter than $1/\gamma$ in the interface channels. The e–e mean free path, 16–25 nm, is well below L_g, as hydrodynamics requires. The momentum mean free path (72–88 nm) is not; Section 5.7 deals with that.

2.3 Semi-analytical coupled-eigenmode theory

Perturbation expansion. The antenna drives the gate–source voltage $\mathrm{Re}[U_0 e^{-i\omega t}]$, with $|U_0|$ small compared with the gate overdrive. Every field $f = n, v, \Theta, u$ of channel i is expanded in this amplitude,

$$f_i = f_{i0} + \mathrm{Re}[F_i e^{-i\omega t}] + f_{i2}, \qquad F = N, V, \Theta_1, U, \quad (4)$$

with first-order phasors ($\mathrm{Im}\, Z > 0$ capacitive) and time-averaged second-order parts. At zero drain bias $v_{i0} = \Theta_{i0} = 0$. The equilibrium under the gate is uniform, so to first order Eqs. (2) become

$$-i\omega N_i + n_{i0} V_i' = 0, \quad (5a)$$

$$(\gamma_i - i\omega) V_i = \frac{q}{m^*} U_i' - \frac{b_i \Theta_{i1}'}{m^* n_{i0}} + \nu_i V_i'', \quad (5b)$$

$$(-i\omega + \tau_E^{-1}) C_{v,i} \Theta_{i1} = -\beta_i V_i', \quad (5c)$$

Friction and viscous heating are second order. Thermal conduction is dropped here although the thermal diffusion length $(\kappa\tau_E/C_v)^{1/2}$ is 80–90 nm in the interface channels (45–50 nm at 1 THz): in the full model, which keeps it, removing it changes V_{ph} at 1 THz by 0.9–2.1%, and halving κ by $\le 0.8\%$. Combined with Eq. (5a), Eq. (5c) gives $\Theta_{i1} = \chi_i N_i / n_{i0}$, with the hot-electron response

$$\chi_i = \frac{\beta_i}{C_{v,i}} \frac{-i\omega}{-i\omega + 1/\tau_E}, \quad (6)$$

Local electrostatics. The gate is much longer than the gate–channel spacing, so per unit area the channel sheets and the gate form the capacitor network of Fig. 1e: C_{ga} (gate to channel a), C_{ab} (through the AlGaN barrier) and C_{b0} (channel b to the buffer). Charge balance on the two sheets,

$$\begin{aligned} C_{ga}(\phi_a - U_g) + C_{ab}(\phi_a - \phi_b) &= -qN_a, \\ C_{ab}(\phi_b - \phi_a) + C_{b0}\phi_b &= -qN_b, \end{aligned} \quad (7)$$

inverts to $\phi_i = c_i U_g - (q/\det)(KN)_i$, $\det = C_{ga}C_{b0} + C_{ab}(C_{ga} + C_{b0})$, with

$$K = \begin{pmatrix} C_{ab} + C_{b0} & C_{ab} \\ C_{ab} & C_{ga} + C_{ab} \end{pmatrix}. \quad (8)$$

Without a back plane, a density wave e^{ikx} sees the buffer capacitance of a half-space, $C_{b0} = \varepsilon_{\mathrm{GaN}}|k|$. It vanishes for the uniform component, for which $c_a = c_b = 1$: the gate voltage enters both channels with unit weight. Adding the degeneracy term of u (Section 2.2) gives

$$U_i = U_g - \frac{1}{q}\sum_j \Big[\frac{q^2}{\det} K_{ij} + \eta_i' \delta_{ij}\Big] N_j. \quad (9)$$

For a single channel the bracket reduces to $q^2/C_g + \eta'$, the classical and quantum (q^2/η') capacitances in series.

Eigenmodes. Substituting Eq. (9) and $\Theta_{i1} = \chi_i N_i / n_{i0}$ into Eq. (5b), and rewriting the viscous term with Eq. (5a) as $\nu_i V_i'' = i\omega\nu_i N_i'/n_{i0}$, makes all three forces proportional to N':

$$\begin{aligned} V_i &= -\frac{(G\,N')_i}{m_i^*(\gamma_i - i\omega)}, \\ G &= \frac{q^2}{\det} K + \mathrm{diag}\Big(\eta_i' + \frac{b_i\chi_i}{n_{i0}^2} - \frac{i\omega m_i^* \nu_i}{n_{i0}}\Big). \end{aligned} \quad (10)$$

Differentiating and applying Eq. (5a) again, $V_i' = i\omega N_i / n_{i0}$, removes the velocities:

$$\begin{aligned} N'' &= -MN, \\ M &= G^{-1}\mathrm{diag}\Big(\frac{i\omega(\gamma_i - i\omega)m_i^*}{n_{i0}}\Big). \end{aligned} \quad (11)$$

Friction, viscosity and the complex χ_i make M non-Hermitian. With its eigenpairs $M\psi_j = k_j^2 \psi_j$ ($\mathrm{Re}\, k_j > 0$) the general solution is

$$N_i(x) = \sum_{j=1}^{2} \psi_{ji}\,[\mathcal{A}_j \cos k_j x + \mathcal{B}_j \sin k_j x], \quad (12)$$

and V_i, U_i follow from Eqs. (10) and (9). Because $C_{b0,j} = \varepsilon_{\mathrm{GaN}}|k_j|$ depends on the mode, each eigenpair is iterated to self-consistency. For a single channel Eq. (11) is the damped gated-plasmon dispersion $k^2 = m^*\omega(\omega + i\gamma)/(n_0 G)$, which becomes $s = \omega/k = (q^2 n_0/m^* C_g)^{1/2}$ when degeneracy, friction and heat are dropped. At 1 THz AlN-cap has two modes (Fig. 3) [9, 10]: an antiphase (179°), channel-a-dominated, strongly damped acoustic-like mode ($|k_1 L_g| = 5.9$) and a channel-b-dominated optical-like mode ($|k_2 L_g| = 0.90$, 47°). Channel a by itself is overdamped

$(\omega/\gamma_a = 0.04)$, so the acoustic-like mode is essentially a diffusive (RC) mode of channel a coupled to the channel-b plasmon.

Access regions and boundary conditions. The access regions and the shared contacts fix the coefficients $\mathcal{A}_j, \mathcal{B}_j$ and the floating drain potential V_d. In the access regions the channels are ungated and the density varies. Treating the electrostatics as local in the channel potentials, with the gate and drain fringing fields as a source, gives $U - u_{\text{ext}} = -qSN$. With the currents $I_i = -qWn_{i0}V_i$, Eqs. (5b) and (5a) then turn each access region into a non-uniform coupled transmission line with one conductor per channel,

$$u' = -\mathcal{Z}I, \qquad I' = -\mathcal{Y}(u - u_{\text{ext}}), \tag{13}$$

with series impedance $\mathcal{Z}_i = m^*(\gamma_i - i\omega)/(q^2Wn_{i0})$ (Drude resistance plus kinetic inductance) and shunt admittance $\mathcal{Y} = -i\omega WS^{-1}$, $S = C_{\text{loc}}^{-1} + \operatorname{diag}(\eta_i'/q^2 + b_i\chi_i/q^2n_{i0}^2)$. The local capacitance $C_{\text{loc}}(x)$ (charge per unit uniform channel potential) and the source $u_{\text{ext}}(x)$, which reproduces the charge the gate and drain fringing fields induce in the grounded sheet, both come from the region's 2D Maxwell capacitance matrix (Section 3). We cascade the line analytically over 10 nm segments with exact cosh/sinh transfer matrices. The cascade starts from $u_i = 0$ at the source and runs to the gate edge. There u_i and I_i match the modal solution of Eqs. (9)–(12) (without the viscous part of U_i), and the cascade continues to the floating drain, $u_i = V_d$, whose electrode the channel currents charge. That gives $2n_c + 1$ linear conditions for the $2n_c$ coefficients and V_d (n_c channels). Because the channels share the contact potentials, no ad hoc condition on $N_b(0)$ is needed, and AC current can pass between them through the drain contact. As in the full model, the gate current follows from the gate charge by reciprocity, and $Z_{dev} = U_g/I_g$.

Rectification. To second order, the time-averaged continuity equation gives $\partial_x(n_{i0}v_{i2} + \frac{1}{2}\operatorname{Re}[N_i\overline{V_i}]) = 0$. The DC flux Γ_i is constant and vanishes on open circuit, so $v_{i2} = -\frac{1}{2}\operatorname{Re}[N_i\overline{V_i}]/n_{i0}$: a DC counterflow cancels the drift current carried by the correlated density and velocity oscillations. In the time-averaged momentum equation we expand the damping as $\langle\gamma v\rangle = \gamma_i v_{i2} + \gamma_{n,i}\frac{1}{2}\operatorname{Re}[N_i\overline{V_i}] + \gamma_{T,i}\frac{1}{2}\operatorname{Re}[\Theta_{i1}\overline{V_i}]$ and the thermal force as $-\langle(b\Theta)'/n\rangle = -b_i\Theta_{i2}'/n_{i0} + (b_i/2n_{i0}^2)\operatorname{Re}[\overline{N_i}\Theta_{i1}']$. The DC force potential then balances the convective, frictional and thermal products:

$$\frac{q}{m^*}u_{i2}' = F_i - \frac{\nu_i}{2n_{i0}}\operatorname{Re}[N_i'\overline{V_i'}] - \nu_i v_{i2}'' + \frac{b_i\Theta_{i2}'}{m^*n_{i0}}, \tag{14}$$

$$F_i = \tfrac{1}{2}\operatorname{Re}[V_i\overline{V_i'}] - \tfrac{1}{2}g_i\operatorname{Re}[N_i\overline{V_i}] + \frac{\gamma_{T,i}\operatorname{Re}[\chi_iN_i\overline{V_i}]}{2n_{i0}} - \frac{b_i\operatorname{Re}[\chi_i\overline{N_i}N_i']}{2m^*n_{i0}^3}, \tag{15}$$

with $g_i = \gamma_i/n_{i0} - \gamma_{n,i}$. The complex χ_i keeps the temperature's phase lead over the density (18° at 1 THz); at 1 THz its imaginary part is about a third of the real part. Integrating Eq. (14) across the gate gives the rectified potential of channel i, $V_{ph,i} = u_{i2}(L_g) - u_{i2}(0)$:

$$V_{ph,i} = \frac{m^*}{q}\Big[\int_0^{L_g}\Big(F_i - \frac{\nu_i}{2n_{i0}}\operatorname{Re}[N_i'\overline{V_i'}]\Big)dx - \nu_i\big[v_{i2}'\big]_0^{L_g}\Big]. \tag{16}$$

The same force, without viscosity, is integrated along both access lines. DC electron heating Θ_2, driven by friction and viscous heating and treated as local (conduction neglected, as above), adds a thermoelectric voltage $q^{-1}\int n_{i0}^{-1}(b_i\Theta_2)'dx$ along the whole channel. The four terms of F_i are convective (Bernoulli) rectification; friction acting on the counterflow v_{i2}, together with the density dependence of the mobility; the hot-electron dependence of the mobility; and the thermal pressure. Without γ_n, γ_T, b and ν, and with the single-channel modes and the Dyakonov–Shur boundary conditions ($U(0) = U_g$, open drain), Eq. (16) reduces exactly to the Dyakonov–Shur result [2]. Treating the $V_{ph,i}$ as Thévenin sources behind their total DC channel resistances R_i (gated plus access), the shared contacts give

$$V_{ph} = \frac{\sum_i V_{ph,i}/R_i}{\sum_i R_i^{-1}}, \tag{17}$$

which allows a DC current to circulate between the channels.

2.4 Antenna, absorbance, responsivity and noise

The antenna, a self-complementary bowtie at the substrate–air interface, is a Thévenin source with $V_{oc} = E_0 l_{\text{eff}}$ and $R_A = Z_w/2$, where $Z_w = Z_{\text{vac}}/\sqrt{\varepsilon_{\text{eff}}}$ ($Z_{\text{vac}} = 120\pi\,\Omega$) and $\varepsilon_{\text{eff}} = (\varepsilon_{\text{sub}} + 1)/2$. It delivers at most $P_{\text{av}} = |V_{oc}|^2/(8R_A)$; with

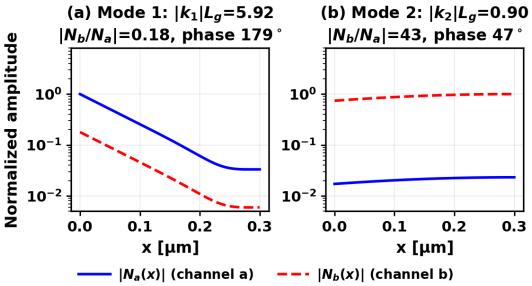


Fig. 3 Density profiles of the two coupled eigenmodes of AlN-cap at 1 THz, on a logarithmic scale, each mode normalized to its larger peak density (blue: channel a; red: channel b, not device colours). (a) Mode 1 (a-dominated), (b) mode 2 (b-dominated).

$R_A = 81.5\,\Omega$ and $l_{\rm eff} = 1.2\lambda_{\rm eff}/\pi$ this is $376\,\mu$W at 1 THz ($l_{\rm eff} = 49.5\,\mu$m). In the quadratic regime R_V and NEP do not depend on E_0. The antenna feeds the load $C_{pad} \parallel (R_c + Z_{dev})$ through a lossless line (Fig. 1d), with $R_c = 0.3\,\Omega$mm$/W$ [35]. By default the line is a quarter-wave transformer rematched at each frequency, $Z_0 = \sqrt{R_A\,\mathrm{Re}\,Z_{\rm load}}$, and $C_{pad} = 0$; Section 5.6 looks at a fixed network (designed at 1 THz, $\theta = \pi f/2f_d$) and at pad capacitances. The absorbance is $A = P_{\rm dev}/P_{\rm av}$, with $P_{\rm dev}$ the power entering Z_{dev}; the responsivity is $R_V = V_{ph}/P_{\rm av}$, and at zero drain bias NEP $= \sqrt{4k_BTR_{\rm out}}/R_V$ (both referred to the available antenna power), where $R_{\rm out}$ is the DC source–drain resistance (channels in parallel plus both contacts). Without a drain current, $1/f$, generation–recombination and shot noise give no voltage noise, so we count only thermal noise. A leaky gate adds the thermal noise of its conductance; thermionic emission and WKB tunneling give gate resistances above $10^{12}\,\Omega$, and even 1 mA/mm at 5 V (250 kΩ) would add $< 0.05\%$.

3 Full-device perturbation model

The semi-analytical theory treats each channel's own electrostatics as local within each region, takes the gate and drain fringing fields as a fixed source, and makes the densities change abruptly at the gate edges. The full model removes these approximations. Each channel is discretized from source to drain on a uniform staggered grid ($\Delta x = 10$ nm; densities and temperatures at nodes, velocities at faces). The potential at channel node j is

$$\phi_j = \sum_k \mathcal{G}_{jk}\lambda_k + g_{g,j}U_g + g_{d,j}V_d, \qquad \lambda_k = -q(n_k - n_{0,k})\Delta x, \tag{18}$$

Here the influence matrix $\mathcal{G}_{jk}$ and the electrode responses g_g, g_d come from a finite-volume solution of the 2D Poisson equation in the (x, z) plane. The gate, source and drain are electrodes, with air above and the layer stack below; the outer boundary has zero normal field (no back plane). Surface-state charge is fixed at THz frequencies. Under a wide gate the solver reproduces ε/t to 10^{-5}. The ohmic contacts act as reservoirs: the density on the contact nodes is fixed, and both channels share each contact's potential. The drain is floating: its charge integrates the arriving channel fluxes, $\dot{Q}_d = -q\sum_i (nv)_{i,\rm last}$, and V_d follows from $Q_d = (C_{dd} + C_{pad,d})V_d + C_{dg}U_g - \sum_k g_{d,k}\lambda_k$ by reciprocity, with the electrode capacitances from the same solution ($C_{dd}W = 1.4$ fF). The antenna drives the gate–source voltage U_g. The equilibrium profile $n_0(x)$ joins the gated and access densities of Section 2.1 with the gate-edge shape of the linear 2D response (Fig. 2b), and γ, ν, η', C_v, β, b and κ are evaluated along it. At the contacts the viscous stress vanishes and $\Theta = 0$.

Equations (2) and (18) are expanded to second order. The first-order system (densities, velocities, temperatures, V_d) is linear and is solved directly for $U_g = 1$; the gate current gives $Z_{dev} = U_g/I_g$, and the antenna chain sets the amplitude U_0. The time-averaged second-order system, for the DC densities, temperatures, $\langle V_d\rangle$ and the circulating flux $\Gamma_a = -\Gamma_b$, is also linear. Its sources are quadratic in the first-order fields: convection $\frac{1}{2}\,\mathrm{Re}[V\overline{\partial_x V}]$, the damping products $-\gamma_0 v_2 - \gamma_n\langle nv\rangle - \gamma_T\langle\Theta v\rangle$, the pressure curvature $\frac{1}{4}\eta''|N|^2$, and the thermal-force, viscous and heating products. The photovoltage is $V_{ph} = \langle V_d\rangle$, the drain relative to the source. Power is conserved to machine precision, and one solution takes about 1 s. Refining Δx from 20 to 5 nm changes V_{ph} at 1 THz by up to 0.4% (Fig. S1b).

The full model reproduces the Dyakonov–Shur limit. We test it for No-cap with $L_g = 10\,\mu$m, $0.1\,\mu$m access regions and 50 GHz, with pressure, viscosity, thermal effects, $\gamma(n)$, $\gamma(T_e)$ and heating switched off. $|V_{ph}|$ then equals $(U_a^2/4\bar{U})f(\omega)$ of Ref. [2] ($\bar{U} = qn_0/C_g$, U_a the gate-to-channel

amplitude at the source edge, f including its finite-length term) within 0.4%, on 20 and 10 nm grids, and also at $L_g = 5\,\mu$m and at 20 GHz.

During this study, Claude Sonnet 5.5 (Anthropic) was used to debug, declutter, and comment custom analysis code, as well as to review the manuscript for paraphrasing, grammar, and consistency. All AI-assisted code was manually tested, and all edited text was reviewed by the authors, who take full responsibility for the accuracy and integrity of the final work.

4 Verification

The nonlinear time-domain solver integrates the same discrete system with the same grid, electrostatic kernel and closures, so it tests the second-order truncation and the implementation, not the physics. The 414 runs cover 375 distinct cases. The gate voltage is ramped, $U_g = [1 - e^{-t/(n_{\mathrm{ramp}}T_p)}]\,|U_0|\cos(\omega t + \arg U_0)$, $n_{\mathrm{ramp}} = 10$, $T_p = 2\pi/\omega$. Each step evaluates V_d and Eq. (18), updates the velocities (explicitly for convection, force, thermal force and optional velocity saturation $\xi_{\mathrm{sat}} = [1 + |v/v_{\mathrm{sat}}|^{1.5}]^{-1/1.5}$ [36]; implicitly for damping and viscosity; upwinding changes V_{ph} by $< 0.2\%$), then the densities, drain charge and Θ. The scheme involves no linearization. The time step is the smallest of $0.5 \cdot 2/\omega_{\max}$ ($\omega_{\max}$: largest linear frequency of the discrete undamped system), $T_p/500$ and, below 1 THz, 2 fs. The scheme's error is first order in Δt (at 0.2 THz, 3.7, 2 and 1 fs give -3.1, -1.9 and -1.2%). The spectral radius of the one-step Jacobian stays below 1 up to 2.2 times this step for the single-channel devices and at least 3 times for AlN-cap (Fig. S1c).

Period averages of V_d converge slowly while the amplitude ramps up (Fig. S1a), so each run advances in chunks of 20 periods and, after $2n_{\mathrm{ramp}}$ periods, stops when successive chunk averages differ by $< 0.5\%$. On grids commensurate with the gate and access lengths (Fig. S1b) both solvers converge with Δx; the model changes by 1.5% from 50 to 5 nm. Their difference stays at 0.6–0.9% on every grid, so it is not a spatial discretization error. Of it, 0.5–0.6 point is the first-order time-step error at 2 fs, and velocity saturation (time domain only) and drive nonlinearity add about 0.1 point each; as $\Delta t \to 0$ the solvers agree within 0.2%. Grids that misplace a region boundary shift V_{ph} by up to 1.2% (7 and 15 nm; 8% at 30–40 nm), not the difference.

Table 2 Verification of the full perturbation model against the nonlinear time-domain solver. Drive 0.3× nominal, except 0.05× below 1 THz, $0.2k_BT/q$ near threshold and $\min(0.3\times, 0.2k_BT/q)$ in the gate-bias runs.

Check	Runs	Max %	Median %
Frequency, 0.2–5 THz	81	3.0	0.7
Trap parameters	114	1.0	0.8
Geometry	111	2.8	0.8
Temperature (77–400 K)	21	0.9	0.8
Energy relaxation time (0.05–2 ps)	24	1.2	0.8
Gate bias ($n_g = 3\times10^{10}$–3×10^{13} cm^{-2})	24	0.9	0.7
Near threshold ($0.2k_BT/q$ drive)	9	1.9	0.6
Grid, $\Delta x = 5$–50 nm	30	0.9	0.8

Table 2 summarizes the checks, and Table S1 lists every run. The verification drive, 0.3× nominal ($E_0 = 3$ kV/m), keeps the response quadratic (Section 5.6). Near threshold it is reduced to $0.2k_BT/q$. Below 1 THz, where the response per volt rises, it is reduced to 0.05× nominal, with the time step held at its 1 THz value (2 fs). Over 0.2–5 THz (29 frequencies per device) the two solvers agree within 3.0% (Fig. 4a). Over the trap, geometry, temperature and bias variations of Section 5 they agree within 2.8%. The six points flanking the sign changes of V_{ph} (No-cap at 3.7–3.9 THz, GaN-cap at 3.3–3.4 THz) are excluded, since the relative deviation is undefined there; the largest, GaN-cap at 3.4 THz (-20%), is where $|V_{ph}|$ is 0.001% of its 0.2 THz value.

5 Results

5.1 Frequency response and absorbance

At 1 THz the full model gives $V_{ph} = -83$, -105 and $-36\,\mu$V for No-cap, GaN-cap and AlN-cap ($R_V = 0.221$, 0.279 and 0.096 V/W; NEP 5.8, 5.1 and 13 nW/$\sqrt{\mathrm{Hz}}$), and -20, -25 and $-7.8\,\mu$V at 2 THz (Fig. 4). The drain stays negative with respect to the source up to 3.3–3.7 THz in the single-channel devices, apart from narrow positive

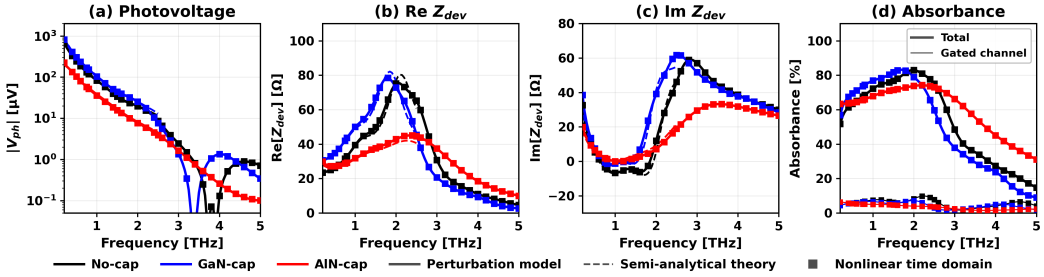


Fig. 4 Frequency response at nominal parameters: (a) photovoltage (squares: nonlinear time-domain solver at 0.05–0.3× nominal drive, rescaled quadratically), (b) real and (c) imaginary part of the device impedance ($e^{-i\omega t}$ convention: $\mathrm{Im}\, Z > 0$ is capacitive), (d) absorbance (thick) and the part absorbed in the gated channel (thin). Solid: perturbation model; dashed: semi-analytical theory (to 2.5 THz, its validated range); squares in (b)–(d): impedance from the time-domain gate current and the absorbance it gives (within 0.6%); small squares in (d): time-domain gated-channel dissipation (within 0.06 points).

windows beyond, and at all frequencies in AlN-cap. The source access region dominates the device impedance, $39-7i$, $50-1i$ and $32+0i\,\Omega$ at 1 THz. The gated channel is capacitively short at THz ($C_g L_g W \approx 20$ fF), and the access region adds its resistance and kinetic inductance, partly offset by its distributed capacitance to the electrodes. The devices absorb 72, 77 and 68% of the available power, but the gated channel takes only 7.5, 6.6 and 5.0%. These values assume ideal matching and a floating drain; a drain pad lowers all three alike (Section 5.6). There is no resonant half-wave feature: at 1 THz $\omega\tau = 1.12$ in the interface channels and 0.04 in channel a, and the drain access region and drain electrode load the drain side.

From 0.5 to 2 THz the semi-analytical theory reproduces the full model's photovoltage within 8.5% (1 THz: +6, +4 and +8%), and up to 2.5 THz its impedance within 10% (Fig. 4a–c; Table S2), including the access-region resonance near 2 THz and the positive $\mathrm{Im}\, Z_{dev}$ above it. Lumped access impedances gave photovoltages up to 55% too low and the wrong sign of $\mathrm{Im}\, Z_{dev}$ above 0.7–1.9 THz. Above 2 THz the gated and access contributions partly cancel and depend on the current near the contacts, which a local line misrepresents; the error grows to 15–68% at 2.5 THz (6–22% with the full 2D capacitance matrix of each access region).

5.2 Interchannel coupling

AlN-cap responds 2.3–2.9 times less than the single-channel devices, and its upper channel is overdamped. To test whether a better upper channel helps, we sweep the channel-a mobility μ_a with the barrier thickness (Fig. 5a). A constant scattering term sets μ_a at nominal density; at $45\,\mathrm{cm^2/Vs}$ it reproduces the nominal R_V within 0.1%. Nowhere in the map (45–1500 $\mathrm{cm^2/Vs}$, 5–50 nm) does the two-channel device reach the single-channel responsivity; the best case (nominal μ_a, 50 nm barrier) reaches 0.46 of it. Thicker barriers improve it only slowly (0.55 at 100 nm). Raising μ_a to $1500\,\mathrm{cm^2/Vs}$ lowers R_V 6.2-fold at 25 nm, to 6% of the single-channel value (2–8% across the map); the time-domain solver confirms the trend within 0.8%.

The responsivity factorizes exactly into absorbance, impedance and the rectification per squared gate voltage,

$$R_V = 2A\, r\, \frac{|Z_{dev}|^2}{\mathrm{Re}\, Z_{dev}}, \qquad r = \frac{V_{ph}}{|U_0|^2}. \tag{19}$$

The loss lies in the impedance, not in the rectification (Fig. 5b): a conductive channel a shunts the source access region, so $\mathrm{Re}\, Z_{dev}$ falls from 31.7 to 7.9 Ω while r changes by less than 25%. The lumped parallel resistance of the two source access channels (15 to 4 Ω) accounts for about half of this. The absorbance halves because the fixed contact resistance ($R_c = 15\,\Omega$) then takes a larger share of the power; without it R_V still falls 3.3-fold. Even for the lossy channel r is about 40% below the single-channel value (full model 2.2 against 3.6–3.8 $\mathrm{mV/V^2}$; semi-analytical 2.5 against 3.9–4.1). Channel a in the access regions causes most of this: with it removed ($N_{TS} = 0.2N_{TS,\mathrm{nom}}$), the full-model r rises to 3.1 $\mathrm{mV/V^2}$ although channel a under the gate is unchanged. A channel a that is harder to contact shunts only slightly less: 10 Ω mm of extra contact resistance on it raises R_V by 28%, to 0.12 V/W (0.56 of the single-channel value; Table S6).

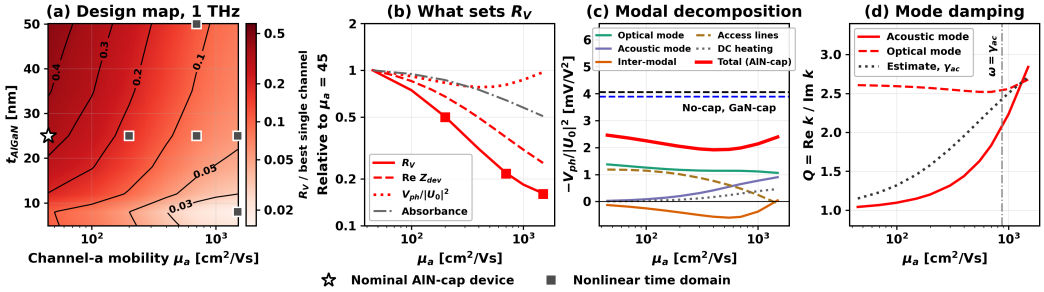


Fig. 5 AlN-cap with the channel-a mobility μ_a as a design variable, 1 THz. (a) Responsivity relative to the better single-channel device at the same barrier thickness (full model; star: nominal; squares: nonlinear time-domain spot checks, within 0.8%). (b) Responsivity and absorbance (full model; squares: nonlinear time-domain solver), device resistance and rectification per squared gate voltage (semi-analytical theory) vs. μ_a at $t_{\mathrm{AlGaN}} = 25$ nm, each relative to its value at 45 cm^2/Vs. (c) Modal decomposition of $-V_{ph}/|U_0|^2$ (semi-analytical theory; component colours are not device colours); dashed horizontal lines: No-cap (black) and GaN-cap (blue). (d) Quality factor $\mathrm{Re}\,k/\mathrm{Im}\,k$ of the two modes and the estimate from γ_{ac}; the vertical line marks $\omega = \gamma_{ac}$.

Because F_i is bilinear in the first-order fields, the gated part of V_{ph} splits exactly into self terms of the optical and acoustic-like (a-dominated) modes and an inter-modal term. The inter-modal term, the sum of the mode pairs (1,2) and (2,1), exists only for coupled channels. The access lines and the DC heating are counted separately (Fig. 5c). At the nominal μ_a the optical mode (channel b's plasmon, $|N_b/N_a| = 43$) and the access lines carry the rectification. As μ_a rises, the optical term changes little (1.4 to 1.1 mV/V^2). The access-line term vanishes (1.2 to −0.1) and is replaced by the acoustic-mode (0.9) and DC-heating (0.5) terms. The inter-modal term opposes the total, by up to 31% near 550 cm^2/Vs, where both modes occupy both channels.

The mode velocities follow from Eq. (10) in the lossless long-wave limit ($C_{b0} \to 0$):

$$s_\pm^2 = \tfrac{1}{2}\left[\mathcal{T} \pm (\mathcal{T}^2 - 4\Delta)^{1/2}\right],$$
$$\mathcal{T} = \mathrm{tr}\,\Sigma, \quad \Delta = \det\Sigma, \quad \Sigma = \mathrm{diag}\left(\frac{n_{i0}}{m_i^*}\right) G_0,$$
$$G_0 = q^2 \begin{pmatrix} C_{ga}^{-1} & C_{ga}^{-1} \\ C_{ga}^{-1} & C_{ga}^{-1} + C_{ab}^{-1} \end{pmatrix} + \mathrm{diag}(\eta_i'), \tag{20}$$

which gives $s_+ = 2.66$ (optical) and $s_- = 1.51 \times 10^6$ m/s (acoustic, set mainly by C_{ab}) for AlN-cap; the adiabatic thermal pressure changes them by $< 0.1\%$. A strictly antiphase mode would have the averaged rate $\gamma_{ac} \approx (n_{b0}\gamma_a + n_{a0}\gamma_b)/(n_{a0} + n_{b0})$ and propagate at 1 THz for $\mu_a \gtrsim 870$ cm^2/Vs. The actual mode ($|N_b/N_a| = 0.18$–0.31) has a quality factor of 1.0–2.8 over the map and reaches the $\omega = \gamma$ value (2.4) only near 1100 cm^2/Vs (Fig. 5d), so γ_{ac} is only an approximate measure.

A second channel thus lowers the zero-bias responsivity mainly through its access regions. Two changes recover part of this loss. Emptying channel a in the access regions ($N_{TS} \le 0.2 N_{TS,\mathrm{nom}}$, Section 5.5) removes the shunt and raises R_V to 0.15 V/W. Gate bias instead depletes channel a under the gate alone, so Z_{dev} and A stay fixed while the full-model r rises from 2.2 to 3.7 mV/V^2. R_V peaks at 0.16 V/W near −4.5 V ($n_a \approx 1.3 \times 10^{12}$ cm^{-2}, NEP 8.3 nW/$\sqrt{\mathrm{Hz}}$), 1.7 times its zero-bias value (Fig. 7d; time-domain solver within 0.8%). About half of this gain is DC heating and most of the rest acoustic-mode and optical rectification of the sparse channel a; $\tau_E = 0.2$–1 ps and removing thermal conduction give 1.5–2.0 times. The two routes do not add: with channel a empty in the access regions, R_V stays at 0.151–0.154 V/W from 0 to −4.8 V, because the sparse gated sheet must stay connected to the contacts through dense access sections. Once the gated section empties (−4.8 V), R_V falls by a third. Neither route reaches the single-channel devices: the upper channel helps only where it is sparse, and a high-mobility upper channel always hurts.

5.3 Physical contributions and hot electrons

Figure 6a adds the physics one term at a time at 1 THz (values in Table S3). Degeneracy and centroid pressure, viscosity and density-dependent mobility each change V_{ph} by $\le 1.2\%$. The adiabatic electron-temperature oscillation does almost nothing through the thermal pressure. Through

hot-electron phonon scattering, $\gamma_T\langle\Theta v\rangle$, it cuts $|V_{ph}|$ by 25% in the single-channel devices and 23% in AlN-cap. Because Θ follows δn, the damping rises where the density is high, and this opposes the density-modulation rectification. DC heating (Θ_2: the thermoelectric voltage and the bolometric mobility change) changes $|V_{ph}|$ by a further -2, -8 and $+0.4$%, so the total hot-electron reduction is 27, 31 and 22%. γ_T depends on the POP matrix element: its effect on $|V_{ph}|$ is 22–27% with ours, 26–29% with the slope of the screened 2D Fröhlich rate (1.0–1.2 times ours) and 9–29% unscreened (0.3–1.4 times), with the device ranking unchanged (Table S6). Over $\tau_E = 0.2$–1 ps, the range reported for GaN channels [28], R_V varies by 5, 17 and 3%. It rises markedly only for $\tau_E \lesssim 0.1$ ps, where Θ no longer follows the density (1.19–1.35 times at 0.05 ps; Fig. 6b). The time-domain solver follows within 1.2% over 0.05–2 ps. The peak temperature swing in the time-domain runs at nominal drive is 5.7 K.

5.4 Gate bias and subthreshold operation

Negative gate bias lowers the gated density n_g (total density at the gate centre, from the Schrödinger–Poisson solution under the gate) and raises the responsivity by three orders of magnitude (Fig. 7a,b); R_{out} rises too, to about 1 kΩ at the NEP minima (Fig. 7b). At 1 THz No-cap reaches $R_V = 235$ V/W at $n_g = 10^{10}$ cm^{-2} and a minimum NEP of 20 pW/$\sqrt{\mathrm{Hz}}$ at $n_g = 1.6 \times 10^{10}$ cm^{-2}; below that density the growing R_{out} takes over; GaN-cap reaches a minimum of 25 pW/$\sqrt{\mathrm{Hz}}$. AlN-cap reaches 30 pW/$\sqrt{\mathrm{Hz}}$ once channel b depletes (near -13 V; channel a: Section 5.2). Near threshold the small-signal regime needs $|U_0| \ll k_BT/q$ (9% below quadratic at k_BT/q for No-cap at 10^{10} cm^{-2}); at $0.2k_BT/q$ the solvers agree within 1.9%. There the plasma phase velocity approaches v_{th} and the gate barrier spans a few mean free paths, so the hydrodynamic minima are approximate (Section 5.7). With the coefficients of Eq. (2c) evaluated at the local (n, T_e), V_{ph} there falls by 8–19% and the No-cap minimum rises to about 22.5 pW/$\sqrt{\mathrm{Hz}}$ (at zero bias by <0.01%; Table S6g). Gate leakage at these biases is negligible in No-cap and GaN-cap (WKB: $< 0.1\,\mu$A); even 1 mA/mm of DC gate current would raise their NEP by only 17–20% through its shot noise. In AlN-cap at -12.7 V, however, field emission through the 5 nm AlN gives about 47 mA/mm and raises the NEP 4.6-fold, so leakage limits its near-threshold operation. Measured AlGaN/GaN detectors reach 25–31 pW/$\sqrt{\mathrm{Hz}}$ at 0.5–0.6 THz [37], with the NEP minimum just above threshold, as here (32 pW/$\sqrt{\mathrm{Hz}}$ at 216 GHz [38]); at 315 GHz our No-cap device gives 6 pW/$\sqrt{\mathrm{Hz}}$ at $R_V = 658$ V/W, though geometry, antenna and calibration differ.

5.5 Temperature, traps and geometry

Cooling from 300 to 77 K raises the zero-bias R_V 1.6, 1.7 and 1.5 times (No-cap, GaN-cap, AlN-cap) and lowers the NEP 4.6, 4.9 and 3.6 times (Fig. 6c). The mobility rises from 1500 to the 7000 cm^2/Vs background and the thermal noise falls as $\sqrt{T}$. At $n_g = 10^{12}$ cm^{-2} the trend reverses (2.1–9.9 V/W, largest near 300 K): near threshold the k_BT/q scale of the density modulation, not damping, limits the rectification.

Surface donors affect only the access density (Fig. S2a,b). At $N_{TS} = 0.1\times$ nominal (No-cap access density 2.6-fold lower) $|V_{ph}|$ of the single-channel devices changes by $\leq$ 9%, and deeper donors (1.0–1.8 eV) by ≤ 3%. In AlN-cap $N_{TS} \leq 0.2\times$ depletes channel a in the access regions, which removes its shunt and raises $|V_{ph}|$ 1.6–1.9-fold. A five-fold buffer-trap density raises $|V_{ph}|$ by 36–38% (single-channel) and 23% (AlN-cap) via the depletion charge below channel b (Fig. S2c,d); a 100-fold barrier-trap density lowers it by up to 16% (No-cap) and 32% (GaN-cap).

The barrier thickness sets the density as well as the coupling (Fig. S3). Thinning it from 25 to 5 nm lowers the single-channel density to 0.14–0.25 $\times$ 10^{13} cm^{-2} and raises $|V_{ph}|$ 2.6–3.4-fold (3.9-fold at 7 nm for GaN-cap); in AlN-cap it mainly shifts density between the channels (-15 to $+4$% over 5–100 nm). Gates below the nominal 0.3 μm change $|V_{ph}|$ by up to 8%, longer ones raise it by up to 20–30% near 0.7–0.9 μm. The source access length, which sets the impedance, matters more: from $L_{sg} = 0.3$ to 2 μm R_V falls 2.1–2.2-fold (No-cap 0.32 to 0.15 V/W), while AlN-cap stays at 0.43–0.47 of No-cap; $L_{gd} = 1$–$3\,\mu$m changes R_V by $\leq$ 1.6% (Table S6). A thicker GaN cap raises

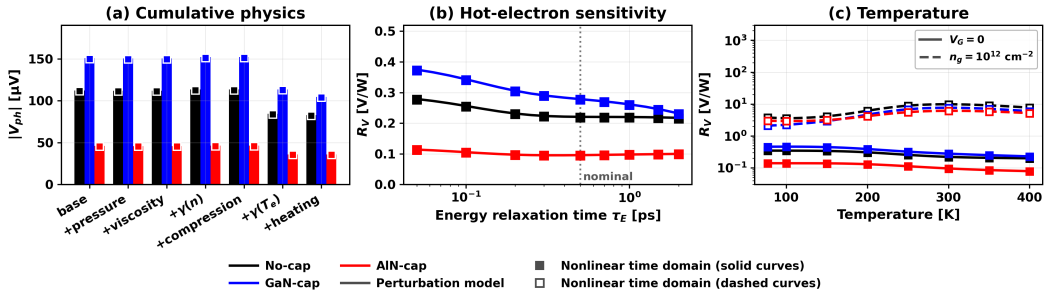


Fig. 6 (a) $|V_{ph}|$ at 1 THz as the physics is added cumulatively (base: isothermal, constant friction; compression: adiabatic temperature oscillation; squares: nonlinear time-domain solver with the same physics, within 0.9%). (b) R_V vs. energy relaxation time (squares: nonlinear time-domain solver; dotted: nominal). (c) R_V vs. temperature at zero bias (solid) and at the bias giving $n_g = 10^{12}$ cm^{-2} (dashed); squares: nonlinear time-domain solver (filled on solid, open on dashed curves; within 0.8% on the dashed curves).

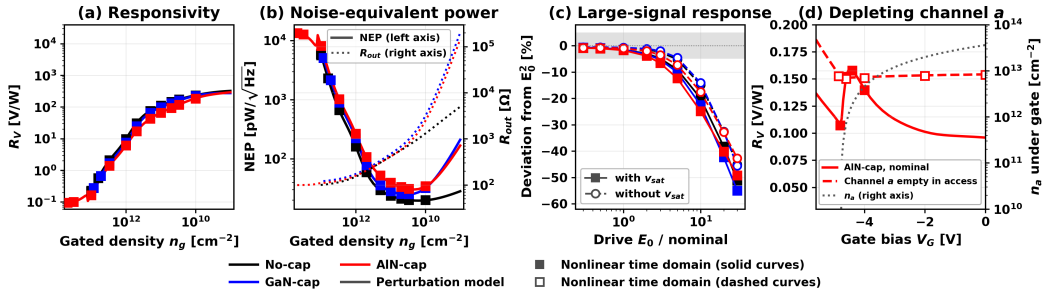


Fig. 7 (a) Voltage responsivity and (b) thermal-noise NEP at 1 THz vs. total gated density (decreasing to the right; the AlN-cap feature near 3×10^{13} cm^{-2} is channel-a depletion), with R_{out} (dotted, right axis) in (b); squares in (a) and (b): nonlinear time-domain checks (NEP from the same runs). (c) Deviation of the time-domain photovoltage from quadratic scaling vs. drive at 1 THz, with (solid, filled squares) and without (dashed, open circles) velocity saturation. (d) AlN-cap R_V vs. gate bias as channel a depletes under the gate (solid; dotted: n_a, right axis) and with channel a also empty in the access regions ($N_{TS} = 0.2N_{TS,\mathrm{nom}}$, dashed; its rise beyond −4.8 V is the depletion of channel b); squares: nonlinear time-domain solver (filled on the solid, open on the dashed curve; within 0.7% with channel a empty); n_a below 10^{10} cm^{-2} is off scale.

$|V_{ph}|$ (1.4-fold from 2 to 10 nm); the AlN thickness changes it by $\leq 7\%$ and the buffer thickness (0.3–3 μm), with no back plane, by $\leq 0.5\%$.

5.6 Antenna parasitics and large signal

A fixed matching network designed at 1 THz keeps R_V within 2% of the rematched value from 0.75 to 1.45 THz and within 10% from about 0.4 to 2.9 THz (Fig. S4a). A 2 fF antenna-feed capacitance in parallel lowers R_V by 7–14% (Fig. S4a,b). Drain loading has more effect: an extra 2 fF of drain-pad capacitance lowers R_V by 42% in the single-channel devices (25% in AlN-cap), and above about 5 fF the drain is AC-grounded and R_V levels off at about 40% of the floating value (Fig. S4b). Realistic 10–50 fF pads therefore put all three devices at 36–43% of their floating responsivity; at 20 fF the single-channel devices still respond 2.1–2.9 times more than AlN-cap (Table S6). For direct or conjugate antenna coupling the ratio is 2.3–3.4 at 0.5–2 THz (Table S6f).

At nominal drive ($E_0 = 10^4$ V/m) the time-domain photovoltage deviates from quadratic scaling by -1.7 to -1.5% at 1 THz (Fig. 7c); ±5% is first exceeded at 5, 3 and 3 times the nominal drive (No-cap, GaN-cap, AlN-cap), and 30 times gives −55 to −49%. Velocity saturation, $\xi_{\mathrm{sat}} - 1 = O(|v/v_{\mathrm{sat}}|^{1.5})$, first rectifies at $O(|U_0|^{3.5})$, since $|\cos|^{1.5}\cos$ averages to zero. Switching it off changes V_{ph} by 0.14% at 0.3× and 6–10% at 10× nominal drive.

5.7 Robustness of the physical model

Adding the refinements of Sections 2.1–2.2 in turn to a triangular-well, parabolic-band baseline (Table S4), Schrödinger–Poisson densities and the fitted compressibility raise R_V by 63%, mostly

because excited subbands raise the heat capacity and lower the temperature swing; the others change it by $\leq 27\%$ and keep the ranking. Uncertain AlN-cap inputs have little effect (Fig. S5): the AlN Schottky barrier (1.5–3.0 eV) changes V_{ph} by $\leq 2\%$, half-relaxed AlN strain by 11%.

The mean free path (72–88 nm) is a quarter of L_g, so we replaced the gated-channel density response by the Mermin (relaxation-time Boltzmann) response [39] (degenerate and non-degenerate forms in Table S5). At zero bias (1–2 THz, $L_g = 0.1$–$1\,\mu$m) and 77 K (Knudsen number up to 1.1) the wavenumber changes by $\leq$1.2% and the semi-analytical V_{ph} by $\leq$0.3%, because the plasma speed is 4–5 times v_F. Near threshold ($\omega/k = 0.9$–$1.5\,v_{th}$) the wavenumber changes by 14–23% and V_{ph} falls by 23–29% (4% at $n_g = 10^{12}\,\mathrm{cm}^{-2}$; Table S5). The estimate omits e–e collisions and the barrier's quasi-ballistic resistance, so it marks where kinetic effects matter, not their size.

6 Discussion

In double-graphene-layer detectors the second layer adds a tunneling nonlinearity [14]; here it adds rectification too weak to offset its loading of the access regions.

This does not contradict the reports that more channels help. Daneshmandian et al. computed absorption [16], not rectification. Meng et al. measured 10 kV/W at 315 GHz [8], which includes the antenna coupling and the gate bias; the bias alone changes R_V by three orders of magnitude here (Section 5.4). Their detector also differs from AlN-cap, but its maximum response (-13.1 V) lies between the two transconductance peaks (-11 and -14 V) [8], where only the lower channel conducts under the gate, the depleted-upper-channel case of Section 5.2. At zero drain bias, for every antenna coupling studied, the coupling lowers the response; devices differing only in channel count would test this.

The method is not specific to this device: the n_c-channel theory needs only the capacitance network, $\eta(n)$, $\gamma(n, T_e)$ and the access-region capacitance, and Eq. (19) and the modal split hold for any antenna-coupled rectifier with linear modes.

Measurements would have to test the channel-a mobility, the Schottky barrier and surface donors, the hot-electron rate and τ_E. (With only the Fang–Howard alloy, phonon and remote terms, the channel-a mobility would be $267\,\mathrm{cm}^2/\mathrm{Vs}$, and AlN-cap would give 0.037 V/W.) A resonant double-channel measurement, as on AlGaN/GaN grating gates [40], remains open.

7 Conclusion

We derived a semi-analytical coupled-eigenmode rectification theory and a full-device perturbation model for single- and two-channel AlGaN/GaN THz detectors, verified in the time domain within 3.0%; the theory reproduces the model's impedance within 10% and photovoltage within 8.5%. The single-channel devices respond best at zero bias (0.22–0.28 against 0.10 V/W at 1 THz) for every antenna coupling studied, and coupling does not help: a better upper channel shunts the access regions and cuts the responsivity 6.2-fold at $1500\,\mathrm{cm}^2/\mathrm{Vs}$. Emptying the upper channel in the access regions, or depleting it under the gate, raises it only 1.6–1.7-fold. Toward threshold the NEP falls to about 20–23 $\mathrm{pW}/\sqrt{\mathrm{Hz}}$ in the hydrodynamic model. For design, keep any upper channel out of the access regions, and sparse or depleted under the gate.

Declarations

- **Funding:** This work was supported by the Indian Space Research Organization (ISRO) under Grant Sanction No. ISRO/RES/3/905.

- **Conflict of interest/Competing interests:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- **Ethics approval and consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Use of AI tools:** Claude Sonnet 5.5 (Anthropic) assisted with code debugging and decluttering, spelling/grammar checks, paraphrasing and proof-reading. The authors verified all results and take full responsibility.
- **Data and code availability:** Python code reproducing every result accompanies this submission (DOI on publication), with the Supplementary Information (Figs. S1–S5, Tables S1–S6).
- **Materials availability:** Not applicable.

- **Author contribution:** S. Ramanujam designed and executed all primary research, developed the methodologies, and drafted the manuscript. Sabareesan. P provided academic advice and reviewed the manuscript text. S. K. Dhayalan provided project tracking and administrative facilitation on behalf of ISRO. M. Varadharajaperumal secured the project funding and provided overall research supervision. All authors reviewed and approved the final manuscript for submission.

References

[1] M. Dyakonov and M. Shur, "Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current," *Phys. Rev. Lett.*, vol. 71, pp. 2465–2468, 1993.

[2] M. Dyakonov and M. Shur, "Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid," *IEEE Trans. Electron Devices*, vol. 43, pp. 380–387, 1996.

[3] W. Knap, Y. Deng, S. Rumyantsev et al., "Resonant detection of subterahertz and terahertz radiation by plasma waves in submicron field-effect transistors," *Appl. Phys. Lett.*, vol. 80, pp. 3433–3435, 2002.

[4] W. Knap et al., "Field effect transistors for terahertz detection: Physics and first imaging applications," *J. Infrared Millim. Terahertz Waves*, vol. 30, pp. 1319–1337, 2009.

[5] A. El Fatimy, S. Boubanga Tombet, F. Teppe et al., "Terahertz detection by GaN/AlGaN transistors," *Electron. Lett.*, vol. 42, pp. 1342–1344, 2006.

[6] J. Singhal, R. Chaudhuri, A. Hickman et al., "Toward AlGaN channel HEMTs on AlN: Polarization-induced 2DEGs in AlN/AlGaN/AlN heterostructures," *APL Mater.*, vol. 10, p. 111120, 2022.

[7] R. M. Chu et al., "AlGaN-GaN double-channel HEMTs," *IEEE Trans. Electron Devices*, vol. 52, pp. 438–446, 2005.

[8] Q. Meng, Q. Lin, F. Han et al., "A Terahertz Detector Based on Double-Channel GaN/AlGaN High Electronic Mobility Transistor," *Materials*, vol. 14, p. 6193, 2021.

[9] L. Wang, X.-S. Chen, W.-D. Hu et al., "The plasmonic resonant absorption in GaN double-channel high electron mobility transistors," *Appl. Phys. Lett.*, vol. 99, p. 063502, 2011.

[10] H. Guo, S. Yang, P. Zhang et al., "Acoustic and optical plasmons excitation in double-channel AlGaN/GaN HEMT under asymmetric boundaries," *AIP Advances*, vol. 14, p. 045224, 2024.

[11] S. Bhardwaj, S. Rajan, and J. L. Volakis, "Analysis of plasma-modes of a gated bilayer system in high electron mobility transistors," *J. Appl. Phys.*, vol. 119, no. 19, 2016, doi:10.1063/1.4950795.

[12] V. V. Popov, O. V. Polischuk, T. V. Teperik et al., "Absorption of terahertz radiation by plasmon modes in a grid-gated double-quantum-well field-effect transistor," *J. Appl. Phys.*, vol. 94, pp. 3556–3562, 2003.

[13] X. G. Peralta, S. J. Allen, M. C. Wanke et al., "Terahertz photoconductivity and plasmon modes in double-quantum-well field-effect transistors," *Appl. Phys. Lett.*, vol. 81, pp. 1627–1629, 2002.

[14] V. Ryzhii, T. Otsuji, M. Ryzhii et al., "Double graphene-layer plasma resonances terahertz detector," *J. Phys. D: Appl. Phys.*, vol. 45, p. 302001, 2012.

[15] Q. Meng, Q. Lin, W. Jing et al., "TCAD Simulation for Nonresonant Terahertz Detector Based on Double-Channel GaN/AlGaN High-Electron-Mobility Transistor," *IEEE Trans. Electron Devices*, vol. 65, pp. 4807–4813, 2018.

[16] F. Daneshmandian, A. Abdipour, and A. N. Askarpour, "Numerical modeling of highly sensitive resonant detection of THz radiation using a multichannel dispersive plasmonic HEMT," *J. Comput. Electron.*, vol. 20, pp. 912–921, 2021.

[17] W. Knap, V. Kachorovskii, Y. Deng et al., "Nonresonant detection of terahertz radiation in field effect transistors," *J. Appl. Phys.*, vol. 91, pp. 9346–9353, 2002.

[18] M. Sakowicz, M. B. Lifshits, O. A. Klimenko et al., "Terahertz responsivity of field effect transistors versus their static channel conductivity and loading effects," *J. Appl. Phys.*, vol. 110, p. 054512, 2011.

[19] A. Lisauskas, U. Pfeiffer, E. Öjefors et al., "Rational design of high-responsivity detectors of terahertz radiation based on distributed self-mixing in silicon field-effect transistors," *J. Appl. Phys.*, vol. 105, p. 114511, 2009.

[20] A. Tomadin and M. Polini, "Theory of the plasma-wave photoresponse of a gated graphene sheet," *Phys. Rev. B*, vol. 88, p. 205426, 2013.

[21] D. A. Bandurin *et al.*, "Resonant terahertz detection using graphene plasmons," *Nat. Commun.*, vol. 9, p. 5392, 2018.

[22] R. Riolo, M. Polini, R. Mannella et al., "Geometric effects in the Dyakonov–Shur theory of terahertz photodetection," *Phys. Rev. B*, vol. 111, p. 235150, 2025.

[23] O. Ambacher, J. Smart, J. R. Shealy et al., "Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 85, pp. 3222–3233, 1999.

[24] J. P. Ibbetson, P. T. Fini, K. D. Ness et al., "Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors," *Appl. Phys. Lett.*, vol. 77, pp. 250–252, 2000.

[25] A. Trellakis, A. T. Galick, A. Pacelli et al., "Iteration scheme for the solution of the two-dimensional Schrödinger–Poisson equations in quantum structures," *J. Appl. Phys.*, vol. 81, pp. 7880–7884, 1997.

[26] O. Ambacher, B. Foutz, J. Smart et al., "Two dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped AlGaN/GaN heterostructures," *J. Appl. Phys.*, vol. 87, pp. 334–344, 2000.

[27] K. Bløtekjær, "Transport equations for electrons in two-valley semiconductors," *IEEE Trans. Electron Devices*, vol. 17, pp. 38–47, 1970.

[28] A. Matulionis, "Hot phonons in GaN channels for HEMTs," *Phys. Status Solidi A*, vol. 203, pp. 2313–2325, 2006.

[29] L. Hsu and W. Walukiewicz, "Electron mobility in $Al_xGa_{1-x}N$/GaN heterostructures," *Phys. Rev. B*, vol. 56, pp. 1520–1528, 1997.

[30] J. H. Davies, *The Physics of Low-Dimensional Semiconductors: An Introduction*, Cambridge University Press, 1998.

[31] G. Bastard, "Energy levels and alloy scattering in InP-In(Ga)As heterojunctions," *Appl. Phys. Lett.*, vol. 43, pp. 591–593, 1983.

[32] P. S. Alekseev, "Negative magnetoresistance in viscous flow of two-dimensional electrons," *Phys. Rev. Lett.*, vol. 117, p. 166601, 2016.

[33] G. F. Giuliani and J. J. Quinn, "Lifetime of a quasiparticle in a two-dimensional electron gas," *Phys. Rev. B*, vol. 26, pp. 4421–4428, 1982.

[34] L. Zheng and S. Das Sarma, "Coulomb scattering lifetime of a two-dimensional electron gas," *Phys. Rev. B*, vol. 53, pp. 9964–9967, 1996.

[35] G. Greco, F. Iucolano, and F. Roccaforte, "Ohmic contacts to Gallium Nitride materials," *Appl. Surf. Sci.*, vol. 383, pp. 324–345, 2016.

[36] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, pp. 2192–2193, 1967.

[37] M. Bauer, A. Rämer, S. A. Chevtchenko et al., "A high-sensitivity AlGaN/GaN HEMT terahertz detector with integrated broadband bow-tie antenna," *IEEE Trans. Terahertz Sci. Technol.*, vol. 9, pp. 430–444, 2019.

[38] J. Sun, Y. Shangguan, X. Li, and H. Qin, "High-dynamic-range terahertz detector based on an antenna-coupled AlGaN/GaN high-electron-mobility transistor," *Opt. Express*, vol. 34, p. 31856, 2026.

[39] N. D. Mermin, "Lindhard dielectric function in the relaxation-time approximation," *Phys. Rev. B*, vol. 1, pp. 2362–2363, 1970.

[40] A. V. Muravjov, D. B. Veksler, V. V. Popov et al., "Temperature dependence of plasmonic terahertz absorption in grating-gate gallium-nitride transistor structures," *Appl. Phys. Lett.*, vol. 96, p. 042105, 2010.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.zip