Effect of Optical Phonon Scattering on the Performance of GaN Transistors

Tian Fang, Ronghua Wang, Huili Xing, *Member, IEEE*, Siddharth Rajan, and Debdeep Jena, *Member, IEEE*

Abstract—A model based on optical phonon scattering is developed to explain peculiarities in the current drive, transconductance, and high-speed behavior of short-gate-length GaN transistors. The model is able to resolve these peculiarities and provides a simple way to explain transistor behavior in any semiconductor material system in which electron–optical-phonon scattering is strong.

Index Terms—f_T, high-electron-mobility transistor (HEMT), optical phonons, saturation current, scattering, transconductance.

I. INTRODUCTION

GALLIUM-nitride-based high-electron-mobility transistors (HEMTs) exhibit a number of features, which have no explanation yet. Polarization induces some of the highest channel 2-D electron gas (2DEG) sheet charge densities in all semiconductors, with values in the range of $n_s \sim 1-4 \times 10^{13}/\mathrm{cm}^2$. If an effective electron velocity of $v_e \sim 10^7$ cm/s is assumed, the drain current per unit device width (W) is expected to be $I_d/W = qn_sv_e \sim 5$ A/mm for $n_s \sim 3 \times 10^{13}/\mathrm{cm}^2$. However, experimentally measured saturation currents fall far short of such values. This indicates that the electron velocity near the source is less than 10^7 cm/s at high V_{gs}. The current-carrying capability of 2DEGs in GaN HEMTs is the first feature to be understood. Second, the "gain" of GaN HEMTs or their transconductance given by $g_m = \partial I_d/\partial V_{\mathrm{gs}}$ shows a sharp drop after reaching a maximum as the gate voltage is increased. Similar behavior has been observed in Si- and InGaAs-based MOSFETs, in which interface roughness scattering limits the carrier transport. In epitaxially grown GaN HEMTs, interface roughness scattering is not the dominant mechanism at room temperature. Third, as expected, the current-gain cutoff frequency (f_T) also has a peak at a certain gate voltage, following the trend of g_m. Earlier explanations attribute these phenomena to a variable source resistance [1]. However, recent results [2] with negligible source access resistances $(\sim 50\ \Omega \cdot \mu m)$ still show the "peaky" behavior of g_m

and f_T versus V_{gs}. The increase of the source (drain) resistance could play a role at high currents but is not the dominant mechanism inducing these features. The purpose of this work is to offer a novel yet simple model of GaN transistor operation that explains all three features. In addition, the model also brings forth a remarkably simple picture of transistor operation when electron–optical-phonon scattering is strong, one that can be used for quantitative predictions.

II. MODEL AND RESULTS

The first hint at resolving the aforementioned features appears upon a close examination of the strength of electron–optical-phonon interactions in GaN. Owing to the light mass of the nitrogen atom (the lightest in all III–V semiconductors), the polar optical phonon energy is high. Combined with the high electronegativity of nitrogen, it results in a very high scattering rate. The mean free path of energetic electrons emitting optical phonons is $\lambda_{\mathrm{op}} \sim a_B^\star \epsilon_\infty/(\epsilon_0 - \epsilon_\infty) \sim 3.5$ nm ($a_B^\star$ is the effective Bohr radius, and ϵ_0 and ϵ_∞ are the static and high-frequency dielectric constants). This is much shorter than that in most other semiconductors (Si and GaAs). It prevents ballistic transport at high fields and damps velocity overshoot effects even for very short gate length GaN devices. Similar high-field phenomena also occur for other high-phonon-energy materials such as carbon nanotubes (CNTs) and graphene [3]–[5]. In CNTs, the quasi-Fermi energy of forward-going and backscattering carriers has a difference of $\hbar\omega_{\mathrm{op}} = 160$ meV, and the current is limited to be $\sim 25\ \mu$A at high bias. In the following, we extend the CNT model to 2DEGs and study the injection velocity in nanoscale GaN HEMTs.

Fig. 1 shows the band diagram of a typical GaN transistor. To understand the characteristics of a short-gate-length device, we inspect the electron distribution at the "source injection point" [6], [7]. The energy and the k-space distribution of 2DEG electrons at this point are depicted for various carrier densities dictated by the gate voltage and the gate-to-source injection plane capacitance. We assume an electrostatically well-designed device with negligible drain-induced barrier lowering and other short-channel effects. Fig. 1(a)–(c) shows the carrier distributions at the injection point. Analogous to a ballistic transistor, the distribution consists of two half Fermi circles, for the right- and left-going carriers, respectively. In a ballistic transistor [6], the difference of the quasi-Fermi levels of the right- and the left-going carriers is the drain bias. However, in GaN transistors, ultrafast optical phonon emission locks this difference at the optical phonon energy ($\hbar\omega_{\mathrm{op}} \sim 92$ meV). As shown in Fig. 1(a), electrons from the highest right-going energy state emit an optical phonon and scatter into the highest empty left-going state. Scattering into the bottom of the subband is Pauli blocked due to the high degeneracy but becomes
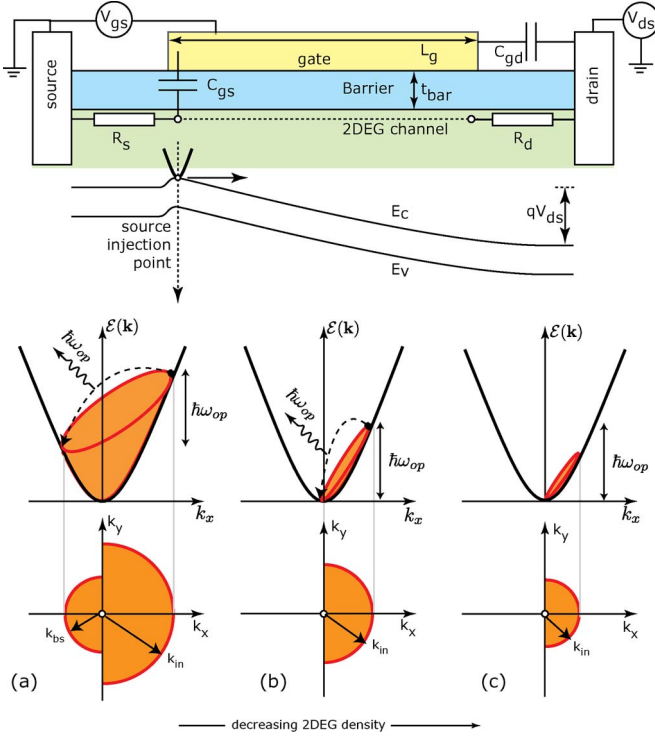
Fig. 1. Transistor cross section, energy band diagram, band structure, and k-space occupation of electrons. k_{in} (k_{bs}) is the quasi-Fermi wave vector of injection (backscattering) carriers.

possible when the condition $k_{in} \leq k_{op}$ is met, as shown in Fig. 1(b). From Fig. 1(b), we get $\hbar^2 k_{op}^2 / 2m^* = \hbar\omega_{op}$, and the crossover 2DEG density is $n_0 = k_{op}^2 / 4\pi \sim 3.8 \times 10^{12}/\text{cm}^2$ at $T = 0$ K (here, $m^* \sim 0.2 m_0$ is the electron effective mass of GaN). For $n_s < n_0$, the highest right-going energy state has an energy of less than $\hbar\omega_{op}$, and optical phonon emission is not possible, so that all the electrons are injected from the source into the channel without any backscattering. At finite temperature, optical phonon emission always happens since the injected carrier distribution has a Fermi–Dirac tail in energy. However, the difference between the quasi-Fermi levels of the injected and backscattered carriers is the optical phonon energy, due to the same reason of optical phonon emission. Now that the carrier distribution at the source injection point in the k-space is identified, the current flowing in the HEMT can be calculated.

The current is found by summing over the group velocities of all occupied states in the k-space; for 2-D carriers in a parabolic band, the result is

$$\mathbf{J} = \frac{\mathbf{J}_{th}}{2\sqrt{\pi}} \left[\mathcal{F}_{1/2}(\eta) - \mathcal{F}_{1/2}\left(\eta - \frac{\hbar\omega_{op}}{k_B T}\right) \right] \quad (1)$$

where $\mathbf{J}_{th} = q v_{th} k_B T m^* / \pi \hbar^2$ is an effective thermal current, with $v_{th} = \sqrt{2 k_B T / m^*}$ being the thermal velocity. $\mathcal{F}_{1/2}(\eta) = (2/\sqrt{\pi}) \int_0^\infty \sqrt{\mu} (1 + e^{\mu - \eta})^{-1} d\mu$ is the Fermi–Dirac integral. The parameter η is $\eta = \mathcal{E}_{in} / k_B T$, where $\mathcal{E}_{in}$ is the injection electron quasi-Fermi level. The two Fermi–Dirac integrals are for the right- and left-going carriers, respectively.

The carrier concentration is the sum of the right- and left-going electrons, given by

$$n_s = \frac{m^* k_B T}{2\pi\hbar^2} \left[\log(1 + e^\eta) + \log\left(1 + e^{\eta - \frac{\hbar\omega_{op}}{k_B T}}\right) \right]. \quad (2)$$
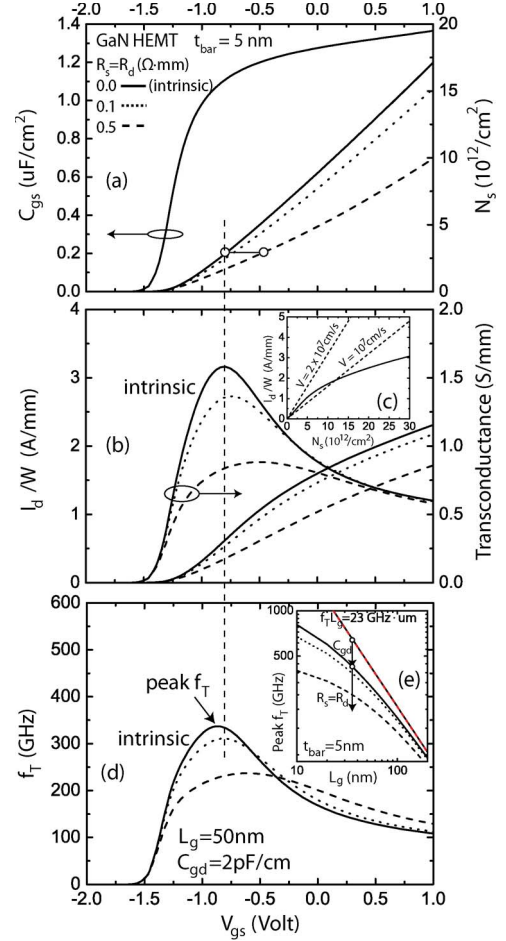

Fig. 2. AlGaN/GaN transistor characteristics calculated using the model. (a) C_{gs} and n_s at the injection point versus the gate voltage. (b) g_m and I_d/W calculated by the optical-phonon-limited model. (c) I_d/W of the optical-phonon-limited model and constant-velocity models. (d) Model result of f_T with gate length $L_g = 50$ nm. (e) Model result of f_T versus L_g.

Equations (1) and (2) give the functional relation between the carrier concentration and saturation current at the injection point. It can be shown from these relations that the current is proportional to the *square root* of the 2DEG density in the high-density (and high-current) regime. The results indicate a significant deviation from a constant-saturation-velocity model. The difference is highlighted in Fig. 2(c).

The transconductance in siemens per meter is given by $g_m = \partial J / \partial V_g = (\partial J / q \partial n_s) C_{gs}$, where C_{gs} is the intrinsic capacitance per unit area between the gate metal and the source injection point. The transconductance can be written in a familiar form as the product of the gate capacitance and the prefactor indicative of an effective electron ensemble velocity, i.e., $v_{eff} = \partial J / q \partial n_s$. Note that the effective velocity is strongly dependent on the carrier density, and therefore, we are forced to refrain from invoking the concept of a saturation velocity that is a material constant. Furthermore, since saturation current is proportional to $\sqrt{n_s}$ at high carrier concentration, the transconductance will *decrease* as $g_m \sim 1/\sqrt{n_s}$. This clearly explains the g_m drop at high carrier concentration. We also note that the transistor intrinsic channel delay is of the form $\tau_c = L_g / v_{eff}$, so that the decrease of v_{eff} at high carrier concentrations also explains the drop of the transistor's cutoff frequency f_T.

In the following calculations, we include the effect of parasitic capacitance and resistances. The extrinsic

transconductance is $g_m^{\mathrm{ext}} = g_m/(1 + R_s g_m)$ since there is a voltage drop across R_s from the source contact to the source injection point. The total delay of a transistor is $\tau_{\mathrm{tot}} = \tau_c + C_{\mathrm{gd}}/g_m + C_{\mathrm{gd}}(R_s + R_d)$, where C_{gd} is the gate–drain overlap capacitance and R_s and R_d are the source and drain resistances (in ohm millimeters). The unity-current-gain cutoff frequency of the transistor is then calculated as $f_T = 1/2\pi\tau_{\mathrm{tot}}$. All the lumped elements are shown in Fig. 1. The only other parameter necessary to quantitatively investigate particular cases is the gate–channel capacitance C_{gs}. To model a realistic GaN HEMT, we use a self-consistent Poisson–Schrodinger solver to calculate C_{gs} as a function of V_{gs}. The calculation gives an accurate dependence of charge at the source injection point on the gate voltage. The calculation accounts for quantization effects in the channel 2DEG. The device structure chosen for highlighting the phonon saturation model is an $\mathrm{Al}_{0.35}\mathrm{Ga}_{0.65}\mathrm{N}/\mathrm{GaN}$ HEMT with a $t_{\mathrm{bar}} = 5$ nm barrier. Fig. 2 shows the device characteristics calculated using the self-consistent Poisson–Schrodinger results as the input.

The calculated C_{gs} is shown as a function of the gate voltage in Fig. 2(a), indicating a threshold voltage of ~ -1.5 V. This figure also shows the 2DEG density at the source injection point for three different R_s values. This sheet density depends on the potential difference between the gate and the source injection point (V_{gs}^0). For $R_s \neq 0$, the external gate voltage is $V_{\mathrm{gs}} = V_{\mathrm{gs}}^0 + JR_s$, and this is what is plotted in Fig. 2(a). As the source resistance increases, a higher gate voltage is necessary to induce the same 2DEG charge. The specific case for $n_s = n_0$ is highlighted by the line connecting two open circles in Fig. 2(a). At room temperature, $n_0 \sim 3 \times 10^{12}/\mathrm{cm}^2$ as shown in the figure. This is the "crossover" density at the source injection point as identified earlier. The 2DEG density at the source injection point is $n_s \sim 10^{13}/\mathrm{cm}^2$ at $V_{\mathrm{gs}} = 1.0$ V if $R_s = 0.5\ \Omega \cdot$ mm.

The drain current and the transconductance are shown in Fig. 2(b) as a function of the gate bias. The intrinsic transconductance peaks at $g_m \sim 1.6$ S/mm. In addition, the peak of g_m happens at the "crossover" point, where the average carrier velocity attains the maximum value. The carrier-concentration-dependent velocity has been measured in [8], where the peak velocity is found at $n_s \sim 3 \times 10^{12}/\mathrm{cm}^2$, which is consistent with our model. Experimental g_m values as high as 0.8–0.9 S/mm have been reported recently [11]. The phonon model thus predicts that, as R_s in GaN HEMTs is lowered, g_m will become more "peaky." The inset [Fig. 2(c)] shows the saturation drain current as calculated from the phonon model in comparison with constant-saturation-velocity models. The constant-saturation-velocity model overestimates the saturation current, particularly at high carrier density. The phonon model follows an $\sim\sqrt{n_s}$ dependence at high densities (for example, ~ 2.5 A/mm at $n_s \sim 2 \times 10^{13}/\mathrm{cm}^2$) and therefore explains experimentally observed currents in GaN transistors.

The calculated current-gain cutoff frequencies are shown in Fig. 2(d) for $L_g = 50$ nm using $C_{\mathrm{gd}} = 2$ pF/cm. The gate dependence of f_T follows that of g_m, repleting with a maximum above the threshold voltage and dropping at higher gate voltages. Severe degradation of f_T and "smoothening" of the sharp features occur with increasing R_s and R_d. To study the gate length dependence of f_T, we choose the peak f_T for each gate length and plot it against L_g in the inset [Fig. 2(e)] for the $t_{\mathrm{bar}} = 5$ nm transistor. The intrinsic cutoff frequency

$f_0 = 1/2\pi\tau_c$ gives the upper limit of the transistor's f_T. The intrinsic $f_T \cdot L_g$ curve in Fig. 2(e) shows that the product is around 23 GHz $\cdot$ μm, corresponding to a constant electron saturation velocity of $v_{\mathrm{sat}} \sim 1.4 \times 10^7$ cm/s [9]. As the parasitic capacitance C_{gd} and access resistance are considered, the peak f_T drops and degrades severely for a smaller gate length device. Based on the calculation, we need $L_g \leq 20$ nm in order to achieve $f_T \sim 500$ GHz in GaN HEMTs.

III. CONCLUSION

The model presented here is able to explain the features of GaN transistors: the current drive, g_m, and f_T behavior with gate bias. Although the model is based on an assumed electron distribution in the $\mathbf{k}$-space, the analogies with the (successful) ballistic transistor model lend credence to it, and its simplicity makes it a useful analytical design tool. In the model, we did not consider the hot-phonon effect [10]. The limitation of the model as presented is that it assumes single subband occupation, which can be violated for high 2DEG densities. In such a case, the model can be extended by summing the current drives in each subband and incorporating intersubband scattering. Effects such as double-peak features in $g_m - V_{\mathrm{gs}}$ plots are then expected to appear as higher subbands are populated.

REFERENCES

[1] T. Palacios, S. Rajan, A. Chakraborty, S. Heikman, S. Keller, S. P. DenBaars, and U. K. Mishra, "Influence of the dynamic access resistance in the g_m and f_T linearity of AlGaN/GaN HEMTs," *IEEE Trans. Electron Device*, vol. 52, no. 10, pp. 2117–2123, Oct. 2005.

[2] K. Shinohara, D. Regan, A. Corrion, D. Brown, S. Burnham, P. J. Willadsen, I. AlvaradoRodriguez, M. Cunningham, C. Butler, A. Schmitz, S. Kim, B. Holden, D. Chang, V. Lee, A. Ohoka, P. M. Asbeck, and M. Micovic, in *IEDM Tech. Dig.*, 2011, vol. 453.

[3] Z. Yao, C. L. Kane, and C. Dekker, "High-field electrical transport in single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 84, no. 13, pp. 2941–2944, Mar. 2000.

[4] D. Jena, "A theory for the high-field current-carrying capacity of one-dimensional semiconductors," *J. Appl. Phys.*, vol. 105, no. 12, p. 123701, Jun. 2009.

[5] X. Luo, Y. Lee, A. Konar, T. Fang, H. Xing, G. Snider, and D. Jena, in *Proc. IEEE DRC Tech. Dig.*, 2008, vol. 29.

[6] K. Natori, "Ballistic metal–oxide–semiconductor field effect transistor," *J. Appl. Phys.*, vol. 76, no. 8, p. 4879, Oct. 1994.

[7] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361–363, Jul. 1997.

[8] C. H. Oxley, M. J. Uren, A. Coates, and D. G. Hayes, "On the temperature and carrier density dependence of electron saturation velocity in an AlGaN/GaN HEMT," *IEEE Trans. Electron Devices*, vol. 53, no. 3, pp. 565–567, Mar. 2006.

[9] G. H. Jessen, R. C. Fitch, J. K. Gillespie, G. Via, A. Crespo, D. Langley, D. J. Denninghoff, M. Trejo, and E. R. Heller, "Short-channel effect limitations on high-frequency operation of AlGaN/GaN HEMTs for T-gate devices," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2589–2597, Oct. 2007.

[10] A. Matulionis, J. Liberis, I. Matulioniene, M. Ramonas, L. F. Eastman, J. R. Shealy, V. Tilak, and A. Vertiatchikh, "Hot-phonon temperature and lifetime in a biased $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}/\mathrm{GaN}$ channel estimated from noise analysis," *Phys. Rev. B, Condens. Matter Mater. Phys.*, vol. 68, no. 3, p. 035338, Jul. 2003.

[11] R. Wang, P. Saunier, X. Xing, C. Lian, X. Gao, S. Guo, G. Snider, P. Fay, D. Jena, and H. Xing, "Gate-recessed enhancement-mode InAlN/AlN/GaN HEMTs with 1.9-A/mm drain current density and 800-mS/mm transconductance," *IEEE Electron Device Lett.*, vol. 31, no. 12, pp. 1383–1385, Dec. 2010.