Hot phonon effect on electron velocity saturation in GaN: A second look

Jacob Khurgin[a)]
Department of Electrical and Computer Engineering, John Hopkins University, Baltimore, Maryland 21218, USA

Yujie J. Ding
Department of Electrical and Computer Engineering, Lehigh University, Betlehem, Pennsylvania 18015, USA

Debdeep Jena
Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA

(Received 1 October 2007; accepted 26 November 2007; published online 17 December 2007)

A theoretical model is developed for electron velocity saturation in high power GaN transistors. It is shown that electron velocity at high electric fields is reduced due to heating of electron gas since the high density of nonequilibrium LO phonons cannot efficiently transfer heat to the lattice. However, the resulting degradation of electron velocity is found to be weaker than previously reported. The results are compared with experimental data, and the ways to improve the efficiency of cooling the electron gas to increase the drift velocity are discussed. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2824872]

GaN is a wide band gap material characterized by a parabolic lowest conduction band separated by ~1.4 eV from the nearest satellite valley. It is also characterized by a high energy of optical phonons ($\hbar\omega_{LO} \approx 92$ meV). These properties make GaN an ideal candidate for high electron mobility transistors (HEMTs) capable of operating at high powers and elevated temperatures.[1] Rule-of-thumb considerations[2] predict that electron velocity saturation in GaN should occur near the onset of strong emission of LO phonons, i.e., when the kinetic energy of electrons becomes comparable to $\hbar\omega_{LO}$. The saturation velocity v_{sat} should then be of the order of $v_0 = \sqrt{\hbar\omega_{LO}/m_e} \approx 3 \times 10^7$ cm/s, where m_e is the electron effective mass. This assumption is supported by the Monte Carlo simulations,[3,4] but the measured electron drift velocity in microwave GaN-based transistors rarely surpasses $v_{sat} \approx 1.5 \times 10^7$ cm/s.

To understand the reasons for this discrepancy, consider distinct properties of GaN separating it from, say, GaAs. In GaAs, the cation and anion do not differ much in masses and electron affinities, while in GaN, they differ by a large amount. A strong polar bond in GaN raises phonon frequencies with a large splitting between LO and TO branches and increases electron scattering by polar optical phonons by at least an order of magnitude in comparison to GaAs. At the same time, the large mismatch between the ion masses in GaN causes substantial splitting between the energies of optical and acoustic branches, such that the zone center LO phonon energy is larger than the highest LA or TA energies by a factor of higher than 2. The decay of LO phonons into two acoustic phonons (the traditional Klemens decay channel[5]) is therefore energy forbidden and the LO phonon lifetime τ_p increases to a few picoseconds. The combination of fast emission of LO phonons and long phonon lifetimes results in a large nonequilibrium LO phonon population. The impact of these "hot" phonons (HPs) on the degradation of the electron velocity saturation is the topic of this work.

Calculations in Ref. 6 predicted that HP can cause very strong degradation of the v_{sat}, with differential mobility becoming essentially zero at fields larger than 100 kV/cm for electron densities in excess of 10^{18} cm^{-3}. This can be easily visualized: as soon as the generation rate of LO phonons exceeds their decay rate τ_p^{-1}, their numbers start increasing avalanchelike. The electron momentum relaxation rate τ_m^{-1} then also sharply increases and the drift velocity v_d gets clamped, i.e., does not change with the field. However, experimental evidence shows that v_d never really gets clamped and continues to increase, albeit with a smaller slope, indicating that τ_m^{-1} does not really increase as dramatically as the hot phonon population.[7] This discrepancy in our view can be attributed to the fact that τ_m^{-1} in Ref. 6 was assumed to be *independent* of the phonon wave vector q. In this work, we perform a more rigorous analysis and show that at elevated electron temperatures T_e, a large fraction of hot phonons are actually "inert" from the point of view of momentum relaxation. The velocity-field curves calculated by taking this into account are closer to experimentally observed values.[7] We show that velocity saturation in GaN is caused not as much by the increased stimulated scattering by the HP but more by the fact that low rate of LO phonon decay causes T_e to rise to the point where conventional spontaneous LO phonon scattering (and to some degree the band nonparabolicity) reduces the drift velocity.

To understand the role of LO phonons, we can write the balance equations for the energy E and drift velocity v_d of an electron as[2,5]

$$\frac{\partial E}{\partial t} = (-eF)v_d - \frac{\hbar\omega_{LO}}{\tau_E(T_e)}, \tag{1}$$

$$\frac{\partial v_d}{\partial t} = \frac{-eF}{m_e} - \frac{v_d}{\tau_m(T_e)}, \tag{2}$$

where F is electric field, τ_E is the energy relaxation time, and τ_m is momentum relaxation time. The steady state solution of Eqs. (1) and (2), $v_d = v_0\sqrt{\tau_m/\tau_E}$ (where $v_0 = \sqrt{\hbar\omega_{LO}/m_e}$), ex-

[a)]Electronic mail: jakek@jhu.edu.

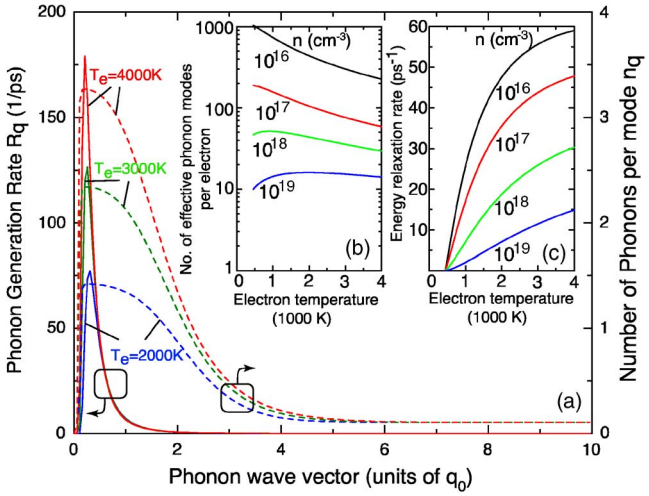

FIG. 1. (Color online) (a) Phonon generation rates (solid lines) and occupation numbers (dotted lines) for different T_e at $N_e=5\times10^{18}$ cm^{-3}. (b) Number of effective phonon modes per electron. (c) Energy relaxation rates for different T_e and N_e.

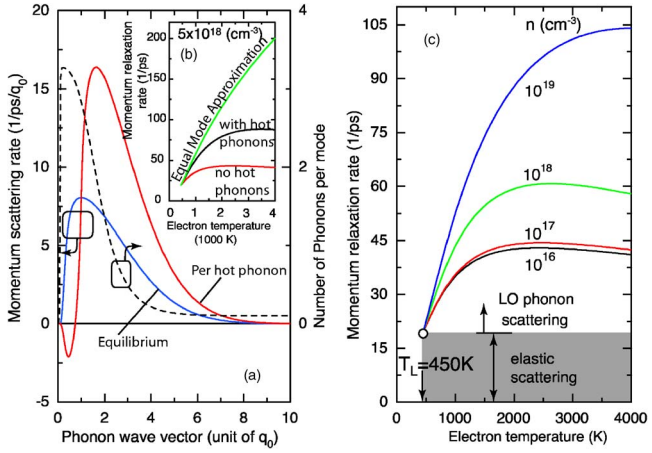

FIG. 2. (Color online) (a) Momentum relaxation rates by equilibrium and hot phonons (solid lines) and occupation numbers (dotted line). (b) Momentum relaxation rate for three different models. (c) Momentum relaxation rates for different T_e and N_e.

emplifies the twofold role of LO phonon scattering. On one hand, LO phonons cause momentum scattering that reduces v_d, but on the other hand, they provide the necessary channels for the energy relaxation of hot electrons thus preventing the electrons from overheating and getting to the high energies in the CB where v_d gets reduced even more. In the most simplistic picture,[2] one can assume $\tau_m \geqslant \tau_E$, which will immediately yield the aforementioned result $v_d \geqslant v_0 \approx 3\times10^7$ cm/s. This assumption indeed holds when the number of phonons is low, and the *spontaneous emission* of phonons is the dominant scattering process. Each time a phonon gets emitted, both the energy and the momentum of the electron get reduced. However, at elevated T_e in the presence of *stimulated* phonon emission processes, the picture gets complicated and needs to be examined in greater detail, since T_e can reach values as high as ten times the lattice temperature, as reported in Ref. 7.

To perform the estimate, we begin with an evaluation of the distribution of LO phonons n_q in the Brillouin zone defined by the balance equation,

$$\frac{dn_q}{dt} = R_q(T_e)[(n_q+1) - n_q e^{\hbar\omega_{\mathrm{LO}}/kT_e}] - \frac{n_q - \overline{n_{qL}}}{\tau_p}, \tag{3}$$

where τ_p is the LO phonon lifetime, $\overline{n_{qL}}$ is the number of phonons at thermal equilibrium with lattice (taken to be at $T_L=450$ K for a biased GaN structure), and $R_q(T_e)$ is the phonon generation rate estimated using standard theory of phonon scattering[5] with all the material parameters of GaN taken from Ref. 3. Here, $R_q(T_e)$ is shown in Fig. 1(a) (solid lines) for different T_e and electron density $N_e=5\times10^{18}$ cm^{-3}. The phonon wave vectors are normalized to the so-called central wave vector $q_0=\sqrt{2m_e\omega_{\mathrm{LO}}/\hbar}$ of an electron with kinetic energy equal to $\hbar\omega_{\mathrm{LO}}$. Clearly, phonon generation occurs only within a tiny fraction of LO modes, but this picture is deceptive because even the tail of the R_q distribution is capable of generating a substantial number of phonons as long as $R_q > \tau_p^{-1}$. Indeed, the distribution of occupation numbers of phonon modes n_q shown by dotted lines in the same figure spreads over a wider range of q, yet the volume of the "active" LO modes still occupies less than 1% of Brillouin zone. This feature can be quantitatively accounted for by introducing an "effective" density of phonon modes interacting with electrons as

$$N_{\mathrm{eff}}(N_e,T_e) = \frac{1}{(2\pi)^3}\int d^3q\left[\frac{n_q - \overline{n_{qL}}}{n_{q,\max} - \overline{n_{qL}}}\right], \tag{4}$$

whereupon N_{eff} is of the order of q_0^3. Thus, less than 10^{20} cm^{-3} out of 5×10^{22} cm^{-3} total LO modes participate in the cooling of the electron gas. It is instructive to plot the number of active LO modes *per electron*, which is the ratio $\eta=N_{\mathrm{eff}}(N_e,T_e)/N_e$, where N_e is the volume density of electrons. As one can see from Fig. 1(b), as N_e and T_e increase, the numbers of "heat dissipation channels" per electron decrease from a thousand to no more than a dozen. It is this bottleneck that reduces the energy relaxation rate $\tau_E^{-1}=\eta\cdot(n_{q,\max}-\overline{n_L})\tau_p^{-1}$ as shown in Fig. 1(c), and leads to overheating.

Next, we consider the momentum scattering rates[5] which we calculate explicitly as functions of q. To better gauge the hot phonon influence, we separate the momentum relaxation rate per phonon into two parts, $R_{\mathrm{mq}}=R_{\mathrm{mq}}^{(\mathrm{eq})}+R_{\mathrm{mq}}^{(\mathrm{hot})}$, where the contribution of equilibrium phonons is $R_{\mathrm{mq}}^{(\mathrm{eq})}=W_{\mathrm{mq}}^{(\mathrm{em})}(\overline{n_{qL}}+1)+W_{\mathrm{mq}}^{(\mathrm{abs})}\overline{n_{qL}} \approx W_{\mathrm{mq}}^{(\mathrm{em})}$ and $R_{\mathrm{mq}}^{(\mathrm{hot})}=(W_{\mathrm{mq}}^{(\mathrm{em})}+W_{\mathrm{mq}}^{(\mathrm{abs})})(n_q-n_{qL}) \approx R_{\mathrm{mq}}^{\prime(\mathrm{hot})}n_q$ is the contribution of HP that includes both emission $W_{\mathrm{mq}}^{(\mathrm{em})}$ and absorption $W_{\mathrm{mq}}^{(\mathrm{abs})}$ of phonons. To account for the relative importance of large q states, we plot in Fig. 2(a) the functions $q^2R_{\mathrm{mq}}^{(\mathrm{eq})}$ and $q^2R_{\mathrm{mq}}^{\prime(\mathrm{hot})}$ for $T_e=4000$ K.

As one can see, the momentum scattering due to equilibrium phonons reaches its maximum in the vicinity of q_0 because it is mostly a *spontaneous* emission process. The scattering by hot phonons, however, reaches its maximum at higher values of q. This has a simple explanation: the absorption of LO phonons with a small q actually can increase the electron momentum, while both absorption and emission of the LO phonon with large q reduce the momentum. If we now plot n_q (dashed curve) in the same figure, we can see that the LO modes that can scatter momentum most effectively are not really "very hot," their temperature is about one-half of T_e, while truly hot modes do not scatter effectively. Thus, unlike the model in Ref. 6 that assumed equal population of all excited phonon modes, τ_m^{-1} is not expected to increase linearly with T_e. This is demonstrated in Fig. 2(b) where we plot the total momentum relaxation rate $\tau_m^{-1}=\tau_{m,0}^{-1}$

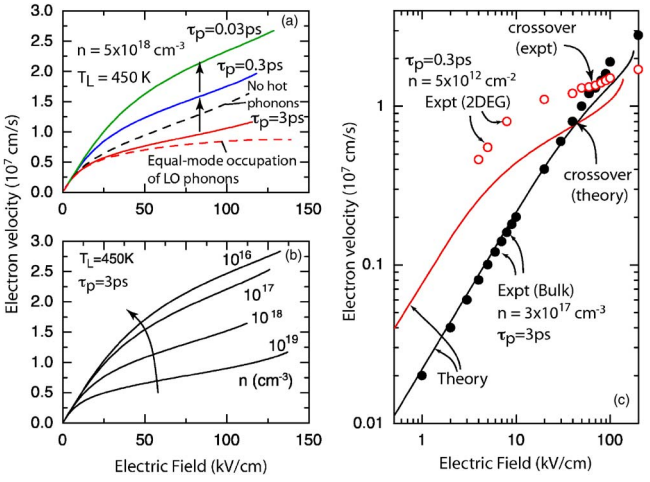

FIG. 3. (Color online) (a) Velocity-field curves using different models and different values of τ_p. (b) Velocity-field curves for various electron densities. (c) Theoretical velocity-field curves (solid lines) and corresponding experimental curves (filled and hollow circles).

$+\tau_{m,\mathrm{eq}}^{-1}+\tau_{m,\mathrm{hot}}^{-1}$ for three cases: without hot phonons, including hot phonons with the correct q distribution, and including hot phonons with the equal-mode occupation approximation as in Ref. 6. Here, $\tau_{m,0}^{-1}=7\ \mathrm{ps}^{-1}$ accounts for all the LO phonon unrelated scattering. Comparing the curves, one can see that at large T_e, the hot phonon contribution is responsible for a twofold increase in the momentum scattering rate. While significant, this increase is *much lower* than the fivefold increase expected under the assumption of equal-mode population. Hence, electron velocity saturation in GaN is expected to be not as severe as predicted in Refs. 6

In Fig. 2(c), τ_m^{-1} for a wide range of electron concentrations are plotted with the correct q distribution of phonons. The concentration-dependent *increase* in τ_m^{-1} caused by hot phonon scattering is by a factor of ~ 2.5 going from $10^{16} \rightarrow 10^{19}\ \mathrm{cm}^{-3}$, while the concentration-dependent *decrease* in τ_E^{-1} caused by the phonon decay bottleneck [Fig. 1(c)] is by a factor of ~ 5. Thus, we can draw the conclusion that the ensuing reduction in v_d is mostly ($\sim 2/3$ of the total) caused by heating due to phonon bottleneck and significantly less ($\sim 1/3$ of the total) by actual HP scattering.

The theoretical velocity-field curves obtained from Eqs. (1) and (2) and the preceding model are shown in Figs. 3(a) and 3(b). As expected from the aforementioned changes in relaxation rates, a significant (factor of 3.5) reduction in v_d occurs as N_e increases from $10^{16} \rightarrow 10^{19}\ \mathrm{cm}^{-3}$. We have also shown in Fig. 3(c) the experimental data[7] obtained in a Si-doped GaN metal-semiconductor field effect transistor sample ($N_e \approx 3\times 10^{17}\ \mathrm{cm}^{-3}$) and two-dimensional AlGaN/GaN HEMT ($N_s \approx 5\times 10^{12}\ \mathrm{cm}^{-2}$), for which we use a different value of $\tau_p \approx 0.3$ ps reported in Ref. 8. Considering the fact that with exception of τ_p, theoretical curves have been obtained without a single adjustable parameter; the agreement is good and the simple model predicts a crossover in the velocity-field curves at a high field, which is observed in the experiment. The theory [Fig. 3(b)] explains why experimental electron velocities in GaN are higher than what was predicted in Ref. 6, as well as the absence of a hard saturation of the electron velocity.

To emphasize the origin of the v_d reduction, we have also plotted in Fig. 3(a) the velocity-field curves for $N_e=5\times 10^{18}\ \mathrm{cm}^{-3}$ obtained under different assumptions. The three solid curves have been obtained using the complete model described above, but while the lowest solid curve uses the realistic $\tau_p=3$ ps, the other two curves are shown to see an improvement (arrows) that can be obtained if one can find a way to reduce the phonon decay time by factors of 10 and 100, respectively. Indeed, removing the phonon decay, bottleneck would raise v_d very close to v_0. The dotted curve, obtained by neglecting the hot phonon component of momentum scattering $\tau_{m,\mathrm{hot}}^{-1}$, shows that while hot phonon scattering is definitely a factor, it cannot alone explain the large decrease of v_d at high N_e. Finally, the lowest dotted curve has been calculated under assumption of equal population of phonon modes and, in full agreement with Ref. 6, predicts clamping of saturation velocity that had not been observed experimentally.

Having identified the phonon decay bottleneck to be the main factor reducing the high field v_d, one can outline possible ways to remove the bottleneck. Actual enhancement of the LO decay into acoustic modes is not feasible unless one can induce changes on the scale of the lattice constant which would amount to growing an entirely new material. What is feasible, however, is to spread the LO phonons around Brillouin zone into region with $q \gg q_0$. This spreading can be attained by the scattering of LO phonons on some random potential caused by either isotopic or alloy disorder. If the disorder occurs on the scale of $q_0^{-1} \approx 10\ \mathrm{nm}^{-1}$ or less, the LO phonons will quickly scatter and occupy significantly larger volume in the reciprocal space. Thus, the all-important ratio $\eta=N_{\mathrm{eff}}/N_e$ and the effective energy relaxation rate will increase. One can alternatively explain this effect as disorder-induced localization of LO phonons, with localized phonons occupying larger fraction of the Brillouin zone and thus providing more channels for transferring the heat from electrons to the lattice. The scattering by localized LO phonons in the presence of disorder needs to be thoroughly investigated before definite conclusions can be drawn, but at the moment, it seems to be a promising direction toward increasing the electron velocity and the speed of high power GaN devices.

This work is supported by DARPA under the direction of Dr. Mark Rosker.

[1] H. Morkoc, *Nitride Semiconductors and Devices* (Springer, Heidelberg, 1999), 1st ed., pp. 39–44.
[2] D. K. Ferry, *Semiconductors* (Macmillan, New York, 1991), 1st ed.
[3] M. Ramonas, A. Matulionis, and L. Rota, Semicond. Sci. Technol. **18**, 118 (2003).
[4] A. Matulionis, J. Liberis, I. Matulioniere, M. Ramonas, L. F. Eastman, J. R. Shealy, V. Tilak, and A. Vertiatchikh, Phys. Rev. B **68**, 035338 (2003).
[5] B. K. Ridley, *Quantum Processes in Semiconductors* (Clanderon, Oxford, 1999), 4th ed., p. 358.
[6] B. K. Ridley, W. J. Schaff, and L. F. Eastman, J. Appl. Phys. **96**, 1499 (2004).
[7] J. Liberis, M. Ramonas, O. Kiprijanovic, A. Matulionis, N. Goel, J. Simon, K. Wang, H. Xing, and D. Jena, Appl. Phys. Lett. **89**, 202117 (2006).
[8] K. T. Tsen, J. G. Kiang, D. K. Ferry, and H. Morkoc, Appl. Phys. Lett. **89**, 112111 (2006).