Hot electron energy relaxation in lattice-matched InAlN/AlN/GaN heterostructures: the sum rules for electron-phonon interactions and hot-phonon effect

J.-Z. Zhang*

School of Physics, Jilin University, Changchun, 130012, China.

A. Dyson

Department of Physics, University of Hull, Hull, HU6 7RX, UK.

B. K. Ridley

School of Computing Science and Electronic Engineering,

University of Essex, Colchester, CO4 3SQ, UK.

(Dated: December 11, 2014)

1

Abstract

Using the dielectric continuum (DC) and three-dimensional phonon (3DP) models, energy relaxation of the hot electrons in the quasi-two-dimensional channel of lattice-matched InAlN/AlN/GaN heterostructures is studied theoretically, taking into account non-equilibrium polar optical phonons, electron degeneracy, and screening from the mobile electrons. The electron power dissipation and energy relaxation time due to both half-space and interface phonons are calculated as functions of the electron temperature T_e using a variety of phonon lifetime values from experiment, and then compared with those evaluated by the 3DP model. Thereby particular attention is paid to examination of the 3DP model to use for the hot-electron relaxation study. The 3DP model yields very close results to the DC model: with no hot phonons or screening the power loss calculated from the 3DP model is 5% smaller than the DC power dissipation, whereas slightly larger 3DP power loss (by less than 4% with a phonon lifetime from 0.1 to 1 ps) is obtained throughout the electron temperature range from room temperature to 2500 K after including both the hot-phonon effect (HPE) and screening. Very close results are obtained also for energy relaxation time with the two phonon models (within a 5% of deviation). However the 3DP model is found to underestimate the HPE by 9%. The Mori-Ando sum rule is restored by which it is proved that the power dissipation values obtained from the DC and 3DP models are in general different in the pure phonon emission process, except when scattering with interface phonons is sufficiently weak, or when the degenerate modes condition is imposed, which is also consistent with Register's scattering rate sum rule. The discrepancy between the DC and 3DP results is found to be caused by how much the high-energy interface phonons contribute to the energy relaxation: their contribution is enhanced in the pure emission process but is dramatically reduced after including the HPE. Our calculation with both phonon models has obtained a great fall in energy relaxation time at low electron temperatures ($T_e < 750$ K) and slow decrease at the high temperatures with the use of decreasing phonon lifetime with T_e. The calculated temperature dependence of the relaxation time and the high-temperature relaxation time $\sim$0.09 ps are in good agreement with experimental results.

PACS numbers: 73.40.Kp, 63.20.kd, 63.22.Np

---

*Electronic address: phyjzzhang@jlu.edu.cn

I. INTRODUCTION

Nitride compound semiconductors such as GaN and AlN which have a wide energy gap withstand high breakdown electric fields and support excellent thermal stability [1]. The heterostructure field-effect transistor (HFET) using a GaN-based heterostructure is very promising for high-power radio frequency and high-mobility operations, owing to its high electron density, high electron drift velocity as well as its high drain breakdown voltage. Particularly interesting are lattice-matched InAlN/AlN/GaN heterostructures [2–5], in which a two-dimensional electron gas forms in the undoped GaN layer near the interface of the GaN layer and the one nanometre thick AlN spacer, due to internal *spontaneous polarization* alone, i.e., with no strain-induced polarization by the piezoelectric effect. As such an electron gas arises in the absence of doping and lattice strain, high electron density ($2 \times 10^{13}/\text{cm}^2$) as well as high drift velocity (3×10^7 cm/s) can occur in the lattice-matched heterostructures. The electrons in the quasi-two-dimensional (quasi-2D) GaN channel are heated up due to high electric power, with the electron temperature T_e elevated above the lattice temperature T_0 (i.e. room temperature). Electron temperatures up to 2500 K have been measured using a microwave noise technique for a lattice-matched InAlN/AlN/GaN heterostructure with an areal electron density of $1.2 \times 10^{13}/\text{cm}^2$ in the GaN channel [2]. Heat dissipation in the GaN channel is a complicated process which includes energy relaxation of the hot electrons mainly by emission of polar-optical phonons, decay of the polar-optical phonons into acoustic phonons via anharmonic interactions and diffusion of the excess acoustic phonons into the remote heat sink. The optical phonon lifetime has been measured for GaN-based heterostructures, which in general falls in the range from 0.1 to 1.7 ps [2, 5, 6] - except for the case of the electron density $8 \times 10^{11}/\text{cm}^2$, and depends on both the electron density [6] and electron temperature [2, 6]. As the optical phonon lifetime is much longer than the Fröhlich scattering time ($\sim$10 fs for GaN [7]), the emission of polar optical phonons is very fast compared to their decay into acoustic phonons, and a large population of nonequilibrium ("hot") phonons are accumulated leading to a slowdown in energy relaxation (termed hot-phonon effect). Hot phonons also result in an increase of electrical resistance and impose limitations on electron drift velocities [5, 8].

The average power dissipation P_d and energy relaxation time τ_E are two key parameters for describing electron energy relaxation in semiconductors under an external electric field.

Apparently both quantities depend on the *electron temperature*, and knowing how they depend on T_e is fundamental to the optimization of the HFET devices [1]. In experimental studies of hot-electron energy relaxation in GaN-based heterostructures [2, 9], the dependence of the electron temperature T_e on the supplied power, which is equal to the power dissipated to the lattice by the hot electrons under steady-state conditions, was *directly* measured using the microwave noise technique, and then the electron energy relaxation time as a function of the electron temperature was deduced by $\tau_E(T_e) = k_B dT_e/dPs$, where P_s is the average power supplied to each electron. For the lattice-matched heterostructures the relaxation time was found to fall sharply at the low electron temperatures and decrease very slowly at the high electron temperatures (>1200 K) [2].

Microscopically, energy relaxation of the hot electrons is governed by scattering with polar-optical phonons. In bulk GaN the polar modes are treated as the longitudinal optical (LO) modes of a single frequency ω_{LO}. In a simplification therefore the polar-optical phonons of the heterostructures are usually taken to be the LO phonons of the bulk material such as GaN. This is referred to as the three-dimensional phonon (3DP) model. According to the dielectric continuum (DC) model, however the eigenmodes of the polar-optical vibrations in the heterostructure include half-space LO modes and interface modes, and all these modes interact with the electrons in the quasi-2D channel. Both phonon models have been used to study energy relaxation and momentum relaxation for GaAs-based quantum wells [10, 11]. For GaN-based quasi-2D systems, the 3DP model was used for an early study on the momentum relaxation and low-field electron transport in GaN quantum wells [12]. In principle, the phonon eigenmodes of quasi-2D GaN heterostructures should be considered for electron transport studies. Indeed, the DC model has been employed in recent years to study electron-phonon scattering associated with the various phonon modes [13], as well as the electron momentum relaxation [14] and energy relaxation [15] in GaN heterostructures. Mori and Ando derived a sum rule [16] from the DC model which showed that the sum of the form factors associated with half-space and interface modes was equal to the form factor for bulk phonons. Later, using a microscopic model Register derived a similar sum rule [17] to prove that the *total* electron-phonon scattering rate in the heterostructure was independent of the phonon basis sets, as long as each set was orthonormal and complete. These sum rules seem to imply that the DC and 3DP phonon models would yield equivalent results for the energy relaxation in GaN heterostructures. However this claim is only partly

4

true, and caution should be taken in applying the sum rules to the energy relaxation in GaN heterostructures. First, in Register's scattering rate sum rule polar scattering with electrons is assumed to be made by phonons of a uniform frequency ω_{LO} [17] (i.e., all the phonon modes are *degenerate*). As is discussed in Appendix A, there is no scattering rate sum rule as the normal modes of the heterostructure including the interface and half-space modes are not degenerate. In fact the interface phonons differ significantly from the half-space phonons in both phonon frequency and electron-phonon coupling strength. Second, as the scalar potential of the interface modes decreases exponentially from the interface according to $e^{-q|z|}$, scattering with interface modes is weak in wide wells making the 3DP evaluation accurate. Indeed, for quantum wells with widths greater than 100 $\mathring{A}$ the 3DP model suffices for the evaluation of scattering rates [11, 18, 19], energy loss rates [20] and momentum relaxation rates [14]. For narrow wells however scattering with interface phonons becomes increasingly important, and as such, the two phonon models yield quite different rates for GaAs quantum wells [19, 20] as well as GaN heterostructures [14]. For GaN heterostructures with a 30 $\mathring{A}$-wide channel the 3DP model underestimates momentum relaxation rates just below the bulk LO phonon energy by 70%, and overestimates rates immediately above the LO phonon energy by 40% compared to the DC model [14]. However as far as we know there has been no comparison of energy relaxation rates for GaN heterostructures based on the two phonon models. Third, in GaN heterostructures screening from the mobile quasi-2D electrons is strong due to the high electron density. The scattering rate sum rule becomes invalid causing quite different DC and 3DP rates when screening is accounted for. Fourth, the scattering rate sum rule is valid with an equilibrium phonon distribution being taken as an important prerequisite [21]. In the GaN HFET where the hot-phonon effect must be taken into account, phonon modes are clearly in nonequilibrium, with the consequence that different modes make different contributions to the energy relaxation process. In this circumstance it is necessary to use the correct normal modes of the GaN heterostructure to calculate the energy relaxation. Previous calculations for GaAs quantum wells have shown that the energy relaxation rates in the hot-phonon regime depend on the phonon models used [20, 22].

Recently, using the DC model the authors calculated hot-electron energy relaxation in a typical lattice-matched InAlN/AlN/GaN heterostructure [15]. We found that the experimentally observed dramatic fall at low T_e [2] was caused chiefly by the fast decreased HPE

and electron screening, while the very slow decrease at the high temperatures was due to the fast optical-phonon decay. In this paper we study the energy relaxation of the hot electrons in the heterostructure with both DC and 3DP models. As the 3DP approach is relatively simple and thus convenient for practical calculation, one of course wants to know how energy relaxation results estimated by this model differ from those calculated with the DC model, and further what causes the discrepancies. We pay particular attention to a quantitative comparison of the T_e-dependencies of the power dissipation and relaxation time at high temperatures calculated with the two models, as the high T_e relaxation process is of great interest in terms of HFET devices. This comparison can be made with regard to only the *total* power loss and relaxation time. On the other hand, there is an advantage of the use of the DC model, in that the contributions from the quasi-2D phonon modes to the energy relaxation can be singled out. It is therefore of great interest to find and understand the behaviours of the half-space modes and interface modes, in particular those of the *interface modes*, in the energy relaxation process in such *narrow* channel GaN heterostructures. The comparison with the 3DP calculations also provides a simple means to examine how the interface phonons contribute. Equally important from the electron gas side is screening. Clearly dynamic effects of screening from the electrons need to be included owing to the high frequencies of the polar modes. For these purposes, a comprehensive study needs to be carried out for the energy relaxation in the GaN heterostructures, in which an emphasis is put on how energy relaxation results from the two phonon models differ when both hot phonons and screening are taken into account. Therefore, using the two phonon models, the power loss and energy relaxation time are calculated as functions of the electron temperature, for a number of phonon scattering processes with electron screening included or excluded. The energy relaxation results from scattering with the half-space and interface phonons are compared and examined, and then the *total* power loss and relaxation time are further compared with those obtained from the 3DP approximation. The sum rules for the electron-phonon interactions are closely checked for the *pure* phonon emission process (i.e., with no HPE). For the *net* phonon emission process where the hot-phonon reabsorption is accounted for, special attention is paid to the difference in the DC and 3DP calculations to examine the 3DP model in the evaluation of the energy relaxation in particular at high electron temperatures. We found that the two models yield very close power loss values and relaxation times, the discrepancies being caused by how much the high energy interface

6

phonons contribute to the energy relaxation. With no hot phonons or screening the power loss calculated from the 3DP model is 5% smaller than the DC power dissipation, whereas slightly larger 3DP power loss (by less than 4% with a phonon lifetime from 0.1 to 1 ps) is obtained throughout the electron temperature range after including both the HPE and screening.

This paper is organized as follows. In Section II, the DC and 3DP models for lattice-matched InAlN/AlN/GaN heterostructures are described where the polar phonon modes and associated electron-phonon interactions are given. Then in Section III a formulation of the power dissipation and energy relaxation time in such heterostructures is presented, taking into account non-equilibrium polar optical phonons, electron degeneracy, and screening from the mobile electrons. Effective numerical techniques in calculating the generation rates and power loss are also described, in terms of handling the integrals involved. In Section IV, first we show results of the non-equilibrium phonon occupation numbers for both half-space and interface modes in a typical lattice-matched InAlN/AlN/GaN heterostructure. These results are used to analyze how hot phonons slow down the quasi-2D electron energy relaxation in the high-temperature region. Then, by choosing two GaN heterostructures with different channel widths we compare power dissipation results from the DC and 3DP models for the simple case with no screening. This is to check the sum rules as well as investigate phonon confinement effects and roles the half-space and interface modes play in the respective pure and net phonon emission processes. In order to examine the usual 3DP approximation in the evaluation of energy relaxation, we further compare the DC and 3DP results of power loss and energy relaxation time in the lattice-matched heterostructure for a number of detailed phonon scattering processes with or without electron screening. Comparisons with the experimental data as well as the bulk GaN situation are also made, and the hot-phonon and screening effects are discussed in great detail. Finally, Section V summarizes the main results obtained. In Appendix A, starting with the detailed generation rate expressions for the half-space, interface and bulk LO phonons, a restoration of the Mori-Ando sum rule is made. Then the form factor sum rule is used to prove that the power dissipation values obtained from the DC and 3DP models are in general different in the pure emission process, except for the limiting case when scattering with interface phonons is sufficiently weak such as in wide GaN channels, or when the degenerate modes approximation is imposed as in the study by Register [17]. These are used to analyze and interpret our energy relaxation results. In

Appendix B, we show the minimum electron kinetic energy for phonon absorption and Fermi-Dirac integrals involved in our energy relaxation calculation, both being functions of the phonon wavevector, which are used for analyzing the non-equilibrium phonon distribution.

II.   ELECTRON-PHONON INTERACTIONS IN THE DC AND 3DP MODELS

There are two types of polar optical modes in a bulk semiconductor of the wurtzite structure such as GaN and AlN owing to anisotropy. The anisotropy however is very small [23], allowing the semiconductor to be treated as a cubic crystal with LO polar vibration modes [8, 24]. The phonon modes of the considered heterostructure $In_xAl_{1-x}N/AlN/GaN$ is dealt with in a simpler way. As the Indium content in the outer barrier is small ($x < 0.2$) the binary alloy $In_xAl_{1-x}N$ is treated as AlN, the same material as the central barrier. This in effect simplifies the lattice matched heterostructure $In_xAl_{1-x}N/AlN/GaN$ as a single heterostructure AlN/GaN [13, 15]. In the dielectric continuum model, the polar vibration modes of a single heterostructure consist of half-space modes and interface modes [16]. The half-space modes have the frequencies of the bulk polar modes of the two constituent materials, whose scalar potentials and electric fields occur in the respective constituent regions. The interface modes have different frequencies from the bulk polar modes, and an interface mode has lattice vibrations and electric fields in both constituent regions.

For the $In_xAl_{1-x}N/AlN/GaN$ heterostructure, let the growth direction be z and the interface between AlN and GaN be at $z = 0$, with the barriers in the space $-L_1 < z < 0$ (where $L_1 = N_1 a$, a being the lattice constant) and the electron-containing active region in the space $0 < z < L_2$ ($L_2 = N_2 a$). Let $\boldsymbol{\rho} = (x, y)$ be the position vector in the plane parallel to the interface. When the electrons are completely confined in the GaN channel the half-space modes only in the GaN region interact with the electrons. Now the half-space modes in the GaN region can be indexed by (q_z, $\mathbf{q}$), all having the LO frequency ω_{LO} of bulk GaN. Here $\mathbf{q}$ is the in-plane phonon wave-vector, and q_z is determined by the fixed end boundary condition imposed on the potential function ($\propto \sin q_z z$), $q_z = l\pi/L_2$ (l=1,2,...,N_2-1). The Hamiltonian of an electron interacting with these half-space modes in the active region can be written as

$$\mathcal{H}_h = \sum_{\mathbf{q}, q_z} \frac{\gamma_{LO}}{\sqrt{2V_2}} \left(\frac{1}{q^2 + q_z^2} \right)^{1/2} e^{i\mathbf{q}\cdot\boldsymbol{\rho}}\, 2\sin q_z z \left[a_{q_z}(\mathbf{q}) + a_{q_z}^+(-\mathbf{q}) \right], \tag{1}$$

where $a_{q_z}(\mathbf{q})$ and $a_{q_z}^+(\mathbf{q})$ are the annihilation and creation operators for the half-space mode of $(q_z, \mathbf{q})$. V_2 is the volume of the GaN half-space $V_2 = AL_2$, with A being the sample area, γ_{LO} is a constant characterizing the electron-LO-phonon coupling strength, and given by $\gamma_{LO}^2 = 2\pi e^2 \hbar \omega_{LO}/\epsilon_{LO}$, where e is the electron charge, and $\frac{1}{\epsilon_{LO}} = (\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0})$, ϵ_0 and ϵ_∞ being the static and high-frequency dielectric constants of bulk GaN. We use cgs units throughout the paper.

The lattice dielectric function of the active GaN material is $\epsilon(\omega) = \epsilon_\infty(\omega^2 - \omega_{LO}^2)/(\omega^2 - \omega_{TO}^2)$, where ω_{TO} is the transverse optical (TO) phonon frequency of bulk GaN. The lattice dielectric function of the simplified barrier AlN is given by $\bar{\epsilon}(\omega) = \bar{\epsilon}_\infty(\omega^2 - \bar{\omega}_{LO}^2)/(\omega^2 - \bar{\omega}_{TO}^2)$, where $\bar{\epsilon}_\infty$, $\bar{\omega}_{LO}$, $\bar{\omega}_{TO}$ are the high-frequency dielectric constant, the LO and TO phonon frequencies of bulk AlN, respectively. Then the frequencies of the interface phonons are determined by $\epsilon(\omega) + \bar{\epsilon}(\omega) = 0$ [16], which yields two solutions ω_n ($n=1,2$; let $\omega_1 < \omega_2$). This shows that, similar to the half-space phonons the interface phonons have no dispersion; that is, the phonon frequencies do not depend on the phonon wavevector. The interface phonon modes can be simply indexed by $(n, \mathbf{q})$ and the electron-interface-phonon interaction Hamiltonian can be written as

$$\mathcal{H}_i = \sum_{n, \mathbf{q}} \frac{\gamma_n}{\sqrt{2A}} \frac{1}{\sqrt{q}} \, e^{i\mathbf{q} \cdot \boldsymbol{\rho}} \, e^{-q|z|} \, \left[a_n(\mathbf{q}) + a_n^+(-\mathbf{q}) \right]. \tag{2}$$

Here γ_n is the electron-interface-phonon coupling strength, $\gamma_n^2 = 2\pi e^2 \hbar \omega_n/\epsilon_n$, where ϵ_n is given by $\frac{1}{\epsilon_n} = 2/[\beta^{-1}(\omega_n) + \bar{\beta}^{-1}(\omega_n)]$, with $\beta(\omega)$ and $\bar{\beta}(\omega)$ being two dimensionless quantities (thus ϵ_n is dimensionless), $\beta(\omega) = \frac{1}{\epsilon_\infty} \frac{(\omega^2 - \omega_{TO}^2)^2}{\omega^2(\omega_{LO}^2 - \omega_{TO}^2)}$, $\bar{\beta}(\omega) = \frac{1}{\bar{\epsilon}_\infty} \frac{(\omega^2 - \bar{\omega}_{TO}^2)^2}{\omega^2(\bar{\omega}_{LO}^2 - \bar{\omega}_{TO}^2)}$. $a_n(\mathbf{q})$ and $a_n^+(\mathbf{q})$ are the annihilation and creation operators for the interface mode $(n, \mathbf{q})$.

In the three-dimensional phonon model, the phonon modes are simply bulk LO modes, which are normalized to the sample volume of the entire heterostructure, $V = AL$, with $L = L_1 + L_2$ being the dimension of the heterostructure in the growth direction z. The bulk LO modes are indexed by the three-dimensional phonon wavevector $\mathbf{Q} = (\mathbf{q}, Q_z)$, where the wavevector Q_z is given by the usual periodic boundary condition, $Q_z = n2\pi/L$, n being an integer, $-(N_1 + N_2)/2 \leq n < (N_1 + N_2)/2$. The electron-LO-phonon interaction is given by the Fröhlich interaction

$$\mathcal{H}_{LO} = \sum_{\mathbf{Q}} \frac{\gamma_{LO}}{\sqrt{V}} \frac{1}{Q} e^{i\mathbf{Q} \cdot \mathbf{r}} \left[a(\mathbf{Q}) + a^+(-\mathbf{Q}) \right], \tag{3}$$

where $a(\mathbf{Q})$ and $a^+(\mathbf{Q})$ are the annihilation and creation operators of the bulk LO mode $\mathbf{Q}$. We note that in the growth direction the wavevector Q_z differs from the q_z of the half-space modes. The two sets of discrete wavevector values result from the two different types of boundary conditions on the respective mode potential functions [11]. The bulk phonon modes are planewaves with electric potential $\propto e^{iQ_z z}$ whereas the half-space modes are standing waves with potential function $\propto \sin q_z z$.

III.   HOT-ELECTRON POWER DISSIPATION AND ENERGY RELAXATION TIME

Confinement in the growth direction z quantizes the motion of an electron, and for the GaN-based heterostructures, the narrow and shallow confinement allows us to consider only the lowest subband [13] which is densely populated by the electrons. Let $\phi(z)$ be the confinement envelope function corresponding to energy ϵ_g for the lowest electron subband. Then the electron wave-function and energy can be written as $\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{A}}\phi(z)e^{i\mathbf{k}\cdot\boldsymbol{\rho}}$, and $E_{\mathbf{k}} = \epsilon_g + \varepsilon_{\mathbf{k}}$, respectively, where $\mathbf{k}$ is the in-plane electron wavevector, and $\varepsilon_{\mathbf{k}}$ is the electron kinetic energy, $\varepsilon_{\mathbf{k}} = \hbar^2 k^2/(2m^*)$, with m^* being the electron effective mass.

The degenerate statistics of a high density of electrons is described by the Fermi-Dirac distribution function, $f(E) = 1/(1 + e^{(E-E_F)/k_B T_e})$, where E_F is the Fermi energy of the quasi-2D electron gas, which is determined by the areal electron density n_A and temperature T_e. The thermal equilibrium population of the phonons of frequency ω at temperature T is given by the Bose-Einstein distribution function, $N(\omega, T) = 1/(e^{\hbar\omega/k_B T} - 1)$. Knowing the electron-phonon interaction Hamiltonians (Sec. II) the energy relaxation of the hot electrons by scattering with interface and half-space phonons or bulk LO phonons can be calculated by Fermi's golden rule. For a given interface mode $(n,\mathbf{q})$ the *number* of phonons which are generated by the hot electrons per unit time can be written as

$$W_n(\mathbf{q}) = \frac{2\pi}{\hbar}\sum_{\mathbf{k}} |M_{n,\mathbf{q}}(\mathbf{k};\mathbf{k}+\mathbf{q})|^2 \left\{(g_n(\mathbf{q}) + 1)f(E_{\mathbf{k}+\mathbf{q}})[1 - f(E_{\mathbf{k}})] - g_n(\mathbf{q})f(E_{\mathbf{k}})[1 - f(E_{\mathbf{k}+\mathbf{q}})]\right\}$$

$$\times \delta(E_{\mathbf{k}+\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_n), \tag{4}$$

where $M_{n,\mathbf{q}}(\mathbf{k};\mathbf{k}+\mathbf{q}) = \langle\psi_{\mathbf{k}}|\mathcal{H}_i|\psi_{\mathbf{k}+\mathbf{q}}\rangle$ is the interaction matrix element associated with electron states $\mathbf{k}$ and $\mathbf{k}+\mathbf{q}$ and phonon mode $(n,\mathbf{q})$ due to the interface phonon scattering,

and $g_n(\mathbf{q})$ is the non-equilibrium interface phonon occupation number. For clearness the index for electron spin has been absorbed into the electron wavevectors.

When the two identities

$$f(E + \hbar\omega)(1 - f(E)) = [f(E) - f(E + \hbar\omega)]N(\omega, T_e), \tag{5}$$

$$f(E)[1 - f(E + \hbar\omega)] = [f(E) - f(E + \hbar\omega)](N(\omega, T_e) + 1), \tag{6}$$

are used, one finds that Eq. (4) can be transformed into a concise expression

$$W_n(\mathbf{q}) = [N(\omega_n, T_e) - g_n(\mathbf{q})]/\tau_n(\mathbf{q}), \tag{7}$$

where $1/\tau_n(\mathbf{q})$ is referred to as the phonon generation rate [8] and given by

$$\frac{1}{\tau_n(\mathbf{q})} = \frac{2\pi}{\hbar} \sum_{\mathbf{k}} |M_{n,\mathbf{q}}(\mathbf{k}; \mathbf{k} + \mathbf{q})|^2 [f(E_{\mathbf{k}}) - f(E_{\mathbf{k}} + \hbar\omega_n)] \delta(E_{\mathbf{k}+\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_n). \tag{8}$$

Expression (7) has a clear physical meaning; that is, $W_n(\mathbf{q})$ is simply the phonon generation *number* $\Delta N = N(\omega_n, T_e) - g_n(\mathbf{q})$ divided by the phonon generation *time* $\tau_n(\mathbf{q})$ for any particular interface phonon mode $(n, \mathbf{q})$.

The polar optical modes decay via the lattice anharmonicity, and the number of interface phonons that decay per unit time is

$$D_n(\mathbf{q}) = [g_n(\mathbf{q}) - N(\omega_n, T_0)]/\tau_p, \tag{9}$$

where τ_p is the phonon lifetime which is assumed to have the same value for all polar modes [11, 22], and $N(\omega_n, T_0)$ is the thermodynamic equilibrium interface phonon number at the lattice temperature T_0. At steady state then one has $W_n(\mathbf{q}) = D_n(\mathbf{q})$ [15, 25], from which one finds the nonequilibrium interface phonon occupation number

$$g_n(\mathbf{q}) = \frac{\frac{1}{\tau_n(\mathbf{q})}N(\omega_n, T_e) + \frac{1}{\tau_p}N(\omega_n, T_0)}{\frac{1}{\tau_n(\mathbf{q})} + \frac{1}{\tau_p}}. \tag{10}$$

This shows that the hot phonon occupation number depends on the relative magnitude of the phonon generation and decay rates $1/\tau_n(\mathbf{q})$ and $1/\tau_p$, with $g_n(\mathbf{q}) \approx N(\omega_n, T_e)$ when $1/\tau_n(\mathbf{q}) \gg 1/\tau_p$, and $g_n(\mathbf{q}) \approx N(\omega_n, T_0)$ when $1/\tau_n(\mathbf{q}) \ll 1/\tau_p$, the latter case being equivalent to neglecting the hot-phonon effect. As $\lim_{q\to 0} \frac{1}{\tau_n(\mathbf{q})} = \lim_{q\to\infty} \frac{1}{\tau_n(\mathbf{q})} = 0$ (refer to Appendix B), from Eq. (10) then one readily finds $\lim_{q\to 0} g_n(\mathbf{q}) = \lim_{q\to\infty} g_n(\mathbf{q}) = N(\omega_n, T_0)$.

11

Similarly, the *number* of half-space phonons generated by the hot electron gas per unit time can be expressed as

$$W_{q_z}(\mathbf{q}) = [N(\omega_{LO}, T_e) - g_{q_z}(\mathbf{q})]/\tau_{q_z}(\mathbf{q}), \tag{11}$$

where the generation rate is

$$\frac{1}{\tau_{q_z}(\mathbf{q})} = \frac{2\pi}{\hbar} \sum_{\mathbf{k}} |M_{q_z,\mathbf{q}}(\mathbf{k}; \mathbf{k} + \mathbf{q})|^2 \left[f(E_{\mathbf{k}}) - f(E_{\mathbf{k}} + \hbar\omega_{LO}) \right] \delta(E_{\mathbf{k}+\mathbf{q}} - E_{\mathbf{k}} - \hbar\omega_{LO}). \tag{12}$$

Here $M_{q_z,\mathbf{q}}(\mathbf{k}; \mathbf{k} + \mathbf{q}) = \langle \psi_{\mathbf{k}} | \mathcal{H}_h | \psi_{\mathbf{k}+\mathbf{q}} \rangle$ is the interaction matrix element associated with electron states $\mathbf{k}$, $\mathbf{k} + \mathbf{q}$ and phonon mode $(q_z, \mathbf{q})$ due to the half-space phonon scattering, $g_{q_z}(\mathbf{q})$ is the hot half-space phonon occupation number, which is given by Eq. (10) with ω_n being replaced with ω_{LO} and $1/\tau_n(\mathbf{q})$ replaced by the half-space phonon generation rate $1/\tau_{q_z}(\mathbf{q})$.

Therefore, in the DC model the average power dissipated per electron is given by

$$P_d = \frac{1}{n_A A} \sum_{\mathbf{q}} \left\{ \sum_n \hbar\omega_n [N(\omega_n, T_e) - g_n(\mathbf{q})]/\tau_n(\mathbf{q}) + \sum_{q_z > 0} \hbar\omega_{LO}[N(\omega_{LO}, T_e) - g_{q_z}(\mathbf{q})]/\tau_{q_z}(\mathbf{q}) \right\}, \tag{13}$$

the right hand side being a sum of the contributions from both the interface and half-space phonons.

In the 3DP model the average power dissipation per electron is simply given by

$$P_d = \frac{\hbar\omega_{LO}}{n_A A} \sum_{\mathbf{Q}} [N(\omega_{LO}, T_e) - g(\mathbf{Q})]/\tau(\mathbf{Q}), \tag{14}$$

where the generation rate $1/\tau(\mathbf{Q})$ has the same expression as $1/\tau_{q_z}(\mathbf{q})$ above [Eq. (12)] except that the Fröhlich interaction matrix element should be used instead, and $g(\mathbf{Q})$ is the hot bulk LO phonon occupation number, whose expression is given by Eq. (10) after substituting the rate $1/\tau(\mathbf{Q})$ for $1/\tau_n(\mathbf{q})$ and the phonon frequency ω_{LO} for ω_n.

Knowing the power dissipation P_d then a hot-electron energy relaxation time τ_E can be defined in the hydrodynamic model [25] through

$$P_d(T_e) = k_B(T_e - T_0)/\tau_E. \tag{15}$$

In-plane isotropy is used to simplify the energy relaxation calculations. Then the in-plane phonon wave-vector dependent quantities such as the phonon generation times $\tau_n(\mathbf{q})$,

$\tau_{q_z}(\mathbf{q})$, $\tau(\mathbf{Q})$ and phonon occupation numbers $g_n(\mathbf{q})$, $g_{q_z}(\mathbf{q})$, $g(\mathbf{Q})$ are reduced to functions of only the magnitude q. Therefore the summation over wavevector $\mathbf{q}$ [Eqs. (13) and (14)] is converted to a double integral (over q, θ) and then reduces to a single integral over only q. We also note that, as neither the half-space, interface phonons nor bulk LO phonons have dispersion (i.e., the phonon frequencies do not depend on wavevector $\mathbf{q}$), the q-dependence of the hot phonon occupation number is dictated by the variation of the generation rate with q [refer to Eq. (10), for instance for the interface modes].

We now include screening. The electron-phonon interactions are screened by the mobile electrons. The response of the electron plasma to a polar disturbance from the lattice is encapsulated by the dielectric function of the electron gas. For a high temperature electron gas as considered here, the Boltzmann distribution function is used to approach the energy distribution of the hot electrons. Then the Lindhard dielectric function of the quasi-2D electron gas reduces to the following form

$$\epsilon(q,\omega) = 1 + F(q)\frac{\kappa_D}{q}\frac{1}{2a}[Z(y - \frac{1}{2}a) - Z(y + \frac{1}{2}a)], \tag{16}$$

where $Z(s)$ is the plasma dispersion function [26]

$$Z(s) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{e^{-x^2}}{x - s}dx, \tag{17}$$

with s being complex, and κ_D is the two-dimensional Debye screening wavenumber, $\kappa_D = 2\pi n_A e^2/(\epsilon_0 k_B T_e)$. In the $\epsilon(q,\omega)$ expression the real arguments of the plasma dispersion function are determined by the two dimensionless quantities y and a,

$$y = \left(\frac{m^*}{2k_B T_e}\right)^{\frac{1}{2}} \frac{\omega}{q}, \quad a = \left(\frac{\hbar^2 q^2}{2m^* k_B T_e}\right)^{1/2}. \tag{18}$$

In Eq. (16) $F(q)$ is a form factor which accounts for the confinement effect on the electron-electron Coulomb interaction due to the *finite* effective channel width of the heterostructure [11],

$$F(q) = \int dz \int dz' |\phi(z)|^2 |\phi(z')|^2 e^{-q|z-z'|}, \tag{19}$$

where $\phi(z)$ is confinement envelope function of the the lowest subband. This form factor is equal to the form factor which was introduced to describe the electron-bulk-LO-phonon interaction [11, 16], namely $F_B(q)$ given by Eq. (A16).

To simplify calculations the plasma dispersion function [integral expression (17)] is approached by a two-pole padé approximant [27],

$$Z(s) = \frac{i\sqrt{\pi} + (\pi - 2)s}{1 - i\sqrt{\pi}s - (\pi - 2)s^2}. \tag{20}$$

Using the properties of the plasma dispersion function [26], it is found that in the static limit $\omega \to 0$ the dielectric function Eq. (16) reduces to

$$\epsilon(q, 0) = 1 + F(q)\frac{\kappa_D}{q}. \tag{21}$$

This is the familiar Debye screening formula which is used to evaluate screening for high-temperature non-degenerate electron gases [11]. In this study, screening from the mobile electrons is handled by dividing the scattering potential, or equally the electron-phonon interaction matrix elements by the dielectric function of the quasi-2D electron gas. The polar disturbance is of course not static and occurs at the *finite* frequency of a particular phonon mode, for instance, a half-space or interface mode. To account for the dynamic effect of screening the frequency ω in the dielectric function is substituted for the corresponding phonon frequency, and the dielectric function is treated as a function of wave-vector q for each finite phonon frequency.

The phonon generation rates $1/\tau_n(q)$, $1/\tau_{q_z}(q)$, $1/\tau(\mathbf{Q})$ are key quantities in the calculation of the power loss and energy relaxation time. The delta function in the rate expressions [Eq. (8), for instance] reflects energy conservation in the scattering of an electron with a phonon of particular mode. The summation over electron wavevector $\mathbf{k}$ is converted to a double integral, which is reduced to a form that is proportional to the difference of two complete Fermi-Dirac integrals of order -1/2 [see Appendix A, Eq. (A1) for $1/\tau_n(q)$]. Accurate evaluation of the integral $F_{-\frac{1}{2}}(x)$ [Eq. (A3)] is important in obtaining the correct energy relaxation results. The calculation should also be efficient as the integration values are input to calculating the generation rates for all interface and half-space or bulk modes in a large phonon wavevector space. To calculate the Fermi-Dirac integral the integrand is transformed to $e^{-x}\frac{e^{\mu}\sqrt{x}}{(1+e^{\mu-x})^2}$ such that the Gauss-Laguerre quadrature technique is used to achieve fast and excellent convergence (25 quadrature points are used). Then these generation rates are inserted into Eqs. (13) and (14) to calculate power dissipation P_d. Again the summation over phonon wavevector $\mathbf{q}$ is transformed to a double integral that is reduced to an integral over only q by in-plane isotropy. Numerical integration is then carried by

14

using the Gauss-Legendre quadrature method, 105 quadrature points being used with the cut-off of q taken to be $8k_0$ (k_0 is the threshold electron wavevector for LO phonon emission, $k_0 = \sqrt{2m^*\omega_{LO}/\hbar}$). To calculate energy relaxation one also needs the Fermi energy E_F, which is determined by the equation $n_A A = \sum_{\mathbf{k}} f(E_{\mathbf{k}})$. Therefore one finds that the Fermi energy is given by $E_F = \epsilon_g + k_B T_e \ln(e^{n_A/n_{T_e}} - 1)$, where n_{T_e} has the dimension of areal density, $n_{T_e} = m^* k_B T_e/(\pi\hbar^2)$.

We model the electron envelope function $\phi(z)$ for the triangular potential well by the Fang-Howard wave-function [28, 29],

$$\phi(z) = \sqrt{\frac{b^3}{2}} z e^{-bz/2}, \tag{22}$$

where b is a variational parameter which is determined by minimizing the total energy of the quasi-2D electron gas. b is related to the areal electron density n_A in the GaN channel via

$$b = \left(\frac{33\pi e^2 m^* n_A}{2\epsilon_0 \hbar^2}\right)^{1/3}. \tag{23}$$

In this wavefunction model an effective channel width d is defined as twice the average penetration depth of the charge in the active GaN region [30, 31]; d is related to the Fang-Howard b parameter via $d = 6/b$.

In this study, the material parameters are taken from Refs. [32–34]. The LO and TO phonon frequencies used for GaN are ω_{LO}=91.13 meV, ω_{TO}=66.08 meV, and for AlN we use ω_{LO}=110.7 meV, ω_{TO}=76.1 meV. The high-frequency dielectric constants are taken to be 5.29 and 4.68 for bulk GaN and AlN respectively. The electron effective mass for GaN is $m^* = 0.22m_0$ (m_0 is the free electron mass), and the lattice temperature is fixed at room temperature 300 K. The optical phonon life-time is a key parameter in the electron energy relaxation study. Thus a range of optical phonon life-time values from 0.1 to 2 ps are taken to examine the hot-phonon effect.

IV. RESULTS AND DISCUSSIONS

A. Non-equilibrium phonon occupation number

The optical phonons contributing to hot-electron energy relaxation are the GaN half-space modes ($\hbar\omega_{LO}$=91.13 meV), and the lower- and higher-energy interface modes ($\hbar\omega_1$=69.70

and $\hbar\omega_2$=102.09 meV, respectively). We first look at the hot-phonon occupancy in phonon wavevector space, as a large number of non-equilibrium phonons of these modes are generated during energy relaxation when their decay is much slower than their emission by the hot electrons. To do this we choose an electron temperature of 1000 K for the electron gas of the areal density of 1.2×10^{13}/cm^2 in a typical lattice-matched InAlN/AlN/GaN heterostructure. The high-T_e experimental value of optical-phonon lifetime 0.1 ps is used for all half-space and interface modes [2]. We calculated the hot-phonon occupation numbers as functions of the in-plane phonon wavevector q for the half-space modes, $q_z = l\pi/L_2$, with mode indices $1 \leq l \leq 100$, as most of these phonons participate in significant electron-phonon scattering. We found that for a given wavevector q the half-space phonon occupation number increases with the mode index l and then decreases after it reaches the maximum value of a certain l_a mode. This is illustrated in Fig. 1(a) where the wavevector q-dependent hot-phonon population are shown for a number of different orders of half-space modes as labelled by the mode indices l. This result can be explained as follows. The hot electrons are confined in a very narrow channel with an effective width of only 44 $\mathring{A}$, whereas the half-space phonons interacting with the electrons are present in the entire GaN region ($0 < z < L_2$). Thus the electron-half-space-phonon overlap integral $\Gamma_H(q_z)$ [Eq. (A7)] as well as the squared interaction matrix element [$\propto \Gamma_H^2(q_z)/(q^2 + q_z^2)$] strongly depends on q_z or equally the mode index l, as displayed in Fig. 1(b). According to Eq. (A6), therefore the q_z-dependence of the squared interaction matrix element dictates the phonon generation rate $1/\tau_{q_z}(q)$ and the variation of the half-space phonon population with q_z. This is quite different from what happens in the usual *square* quantum wells such as GaAs/AlGaAs quantum wells, where the hot-phonon population decrease as the confined-mode index in the growth direction z increases [20]. This is because both the electrons and phonons are confined in the same *well* region [11], resulting in stronger electron-phonon interaction with a larger overlap integral for the lower-order phonon mode than the higher-order mode. We also see from Fig. 1(a) that the half-space phonons of different orders l have peak occupation numbers occurring at different phonon wavevectors q. In the two limits $q \to 0$ and $q \to +\infty$, however the occupation numbers of all order half-space modes approach the common thermal equilibrium value at room temperature, $N(\omega_{LO}, T_0) = 1/(e^{\hbar\omega_{LO}/k_B T_0} - 1) = 0.03$. The interface phonon occupation numbers are shown in Fig. 2(a) for both the lower-frequency (ω_1) and higher-frequency (ω_2) interface modes. The lower-frequency

phonons IF_1 have a smaller population, whilst the higher-frequency modes IF_2 have a higher peak and are densely populated in a broader q-region where the IF_2 phonon generation rate $1/\tau_2(\mathbf{q}) \gg 1/\tau_p$ because they have much greater electron-phonon coupling strength than the IF_1 modes (the Fröhlichlike coupling constants for the interface modes, $\alpha_n = \frac{e^2}{\hbar}(\frac{m^*}{2\hbar\omega_n})^{1/2}/\epsilon_n$ [16], are α_1=0.02 and α_2=0.5; that is, the higher frequency IF$_2$ phonons have more than one order of magnitude larger coupling strength than the lower frequency IF$_1$ phonons). According to Eq. (10), the occupation number $g(q)$ of the frequency ω modes is restricted to the range $N(\omega, T_0) \le g(q) \le N(\omega, T_e)$. Therefore the peak occupation number of the ω_n interface phonons is smaller than $N(\omega_n, T_e)$, namely the Bose-Einstein distribution function at electron temperature T_e. Further, with $\omega_1 < \omega_{LO} < \omega_2$, the minimum occupation number of the half-space modes is larger than that of the higher-frequency IF_2 modes but smaller than the minimum occupation number of the lower-frequency IF_1 modes [comparing Fig. 1(a) and Fig. 2(a)]. For both half-space and interface phonons, the occupation number curves [Figs. 1(a) and 2(a)] have a *steep* edge on the small-wavevector side and a slow slope on the large-q side. We found that this originates from the wavevector q-dependence of the energy Δ_q [that is, the *minimum* electron kinetic energy for phonon absorption, Eq. (A9)], as is shown in Fig. 9(a) and discussed in Appendix B. Further, we found that the q-dependent non-equilibrium phonon occupation number as shown above is governed by the Fermi-Dirac integral $F_{-\frac{1}{2}}(\xi_q)$ [with ξ_q being given by Eq. (A8)] as a function of the phonon wavevector q [Fig. 9(b)], which is discussed in Appendix B.

The influence of screening from electrons on hot-phonon population is shown in Fig. 2(b) for the interface modes. Compared to the non-screening calculation [Fig. 2(a)], static Debye screening has significantly reduced the phonon population, and in particular the population of the IF_1 phonons are reduced substantially as these low-frequency phonons are restricted to only a small-q region ($q \ll k_0$) of wavevector space where screening from the electrons is very strong with large $F(q)/q$ and hence large values of dielectric function $\epsilon(q, 0)$ [Eq. (21)]. Recall that $F(q)$ is the form factor associated with the Coulomb interaction [Eq. (19)], which increases as the wavevector q decreases. When dynamic screening is used [Fig. 2(b), dashed and dotted lines], we see that the interface phonon population becomes narrower in wavevector q-space, with the occupation numbers at small q wavevectors being increased rather than decreased, compared to the case of excluding screening in Fig. 2(a). This anti-screening arises due to the dispersion of the quasi-2D electron plasma oscillation frequency,

namely $\omega_{pl}(q) = \sqrt{2\pi e^2 nq F(q)/(\epsilon_0 m^*)}$, which is smaller than the phonon frequency at long wavelengths [11]. In this circumstance the electron plasma cannot move sufficiently fast to cause screening to the polar disturbance from the lattice [35]. These hot phonon population results will be used to analyze the power dissipation calculation below.

B. Hot electron power dissipation and energy relaxation time

The three-dimensional phonon (3DP) model has been widely used to evaluate electron energy and momentum relaxation rates for quasi-2D semiconductor systems [11, 36–38]. Here we compare hot electron power dissipation in GaN based heterostructures calculated with the DC and 3DP models. First we examine the sum rules as applied for the energy relaxation in GaN heterostructures. We include only the hot phonon effect and do not consider screening. To do this, we consider two heterostructures with different *effective* channel widths, namely a strained $Al_{0.05}Ga_{0.95}N/GaN$ heterostructure [39] with a wide well of 110 $\mathring{A}$ (corresponding to an areal electron density of $7 \times 10^{11}/cm^2$) and a lattice matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure [2] with a narrow well of 44 $\mathring{A}$ (corresponding to electron density $1.2 \times 10^{13}/cm^2$). Then the average power dissipated per electron is calculated as functions of the electron temperature with hot phonons being excluded or included, for the latter case two phonon life-time values being used, τ_p=0.5 and 2 ps, to investigate the HPE. Figs. 3(a) and 3(b) show the results with the electron temperatures ranging from room temperature up to 2500 K. For the strained heterostructure with a wide channel, as shown in Fig. 3(a) the two phonon models yield literally the same power dissipation. In this case, the interface phonon scattering with potential decreasing exponentially according to $e^{-q|z|}$ [Eq. (2)] is very weak, as the average value of the position for electrons $\bar{z}$ which is half the effective channel width [16] is $\bar{z}$=55 $\mathring{A}$, whereas the characteristic wavevector k_0 for the interface phonons is around 0.07 $1/\mathring{A}$, making $k_0\bar{z} \approx 4$. The form factor for the interface phonons [Eq. (A17)] $F_I(q)$ can be neglected, and as is proved in Appendix A, then power dissipation values given by the two phonon models are equal, which is consistent with both sum rules [16, 17]. For the heterostructure with a narrow 44$\mathring{A}$ channel, in contrast, the interface phonon scattering is significantly enhanced, and in this case the two phonon models yield different power loss values (see proof in Appendix A). Therefore a clear difference is seen between the power dissipation curves calculated with the two phonon models [Fig. 3(b)].

18

In the pure emission case the DC model yields higher power dissipation than the 3DP approximation. Taking the HPE into account, however the 3DP power loss becomes larger. To find the cause we need to separate and check contributions to the power dissipation from the half-space modes, the lower and higher frequency interface modes respectively [Fig. 4(a)]. We see that the half-space phonons dominate the energy relaxation process (due to the large density of states of the half-space modes) and contribute larger power dissipation than the interface phonons, whilst the lower-frequency IF$_1$ phonons make only a small contribution due to their low energy, small coupling strength and narrow population distribution in q-space (refer to Fig. 2 and the preceding subsection). Without HPE the power dissipation due to the *high-energy* IF$_2$ phonons increases rapidly with T_e in particular above 1000 K compared to the half-space phonons. For instance, the IF$_2$ power dissipation at T_e=2500 K has risen to 46% the power dissipation due to the half-space phonons. This causes a larger total power dissipation with the DC model than the 3DP model [Fig. 3(b)]. When the hot phonons are taken into account, however we see from Fig. 4(a) that the power dissipation due to the IF$_2$ phonons drops dramatically by about 85%, because the IF$_2$ phonon generation number per unit time $W_2(q)$ [Eq. (7)] is substantially reduced [compare the two curves in Fig. 4(b)], as the nonequilibrium IF$_2$ phonons with a broad population distribution in wavevector space (as illustrated in Fig. 2) are re-absorbed. As a result the total DC power dissipation becomes smaller than that evaluated with the 3DP model [Fig. 3(b)]. Therefore the difference between the DC and 3DP results is due to the IF$_2$ phonons - their contribution to the power loss is enhanced at the high electron temperatures in pure emission but is dramatically reduced after including the HPE. In recent studies on GaN heterostructures, we found that interface phonon absorption causes negative momentum relaxation rates [14], and also an increased interaction with the IF_2 modes leads to a reduction of phonon lifetime [13].

In both heterostructures the calculations with both phonon models [Figs. 3(a) and 3(b)] show that the electron power dissipated increases rapidly with T_e at low temperatures but the increase becomes slower at high temperatures, similar to that which occurs in Si-doped bulk GaN [7]. In the simpler case with no screening or hot phonons, the phonon generation number is $\Delta N(T_e) = N(\omega_{LO}, T_e) - N(\omega_{LO}, T_0)$ for bulk LO modes, and the temperature dependence of the power loss P_d, according to Eq. (14), is determined entirely by the product of the generation number $\Delta N(T_e)$ and the *total* generation rate of all bulk modes $\sum_{\mathbf{Q}} \frac{1}{\tau(\mathbf{Q})}$.

The low-temperature power loss is dictated by the *exponential* increase of $\Delta N(T_e)$ with T_e, $\Delta N(T_e) \approx e^{-\hbar\omega_{LO}/(k_B T_e)} - e^{-\hbar\omega_{LO}/(k_B T_0)}$. At high temperatures (>1200 K), whilst the generation number $\Delta N(T_e)$ increases with T_e the total generation rate is reduced, resulting in the *slow* rise of power dissipation. Physically the generation rate decreases as the difference in electron occupation numbers within the phonon energy $\hbar\omega_{LO}$ becomes smaller at a higher electron temperature.

We now include screening to make a comprehensive study of the electron power dissipation calculated with the DC and 3DP models. When both screening and hot phonons are included, strictly speaking the scattering rate sum rule is not applicable, and then one needs to find how much discrepancy the 3DP evaluation yields with respect to the DC calculation. Thus calculations were performed using each phonon model for a number of cases, namely, (i) excluding hot phonons and screening, (ii) including only the HPE, (iii) including only static (Debye) screening, (iv) including only dynamic screening, (v) including both HPE and static screening, and (vi) including both the HPE and dynamic screening. The results are show in Fig. 5 for the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure (with a 44-$\mathring{A}$-wide channel), where a polar optical phonon lifetime of 1 ps is used for all the cases of including the HPE. Several points can be made by comparing the power dissipation values in the various cases. First, the power dissipation is substantially reduced by static screening [compare cases (i) and (iii)], whereas the reduction using dynamic screening is much smaller, which is only $\sim 30\%$ the reduction caused by Debye screening [compare cases (i), (iii) and (iv)]. When both screening and hot phonons are included, the power loss values obtained with the static and dynamic screening models get closer as T_e increases; at T_e=2500 K the power loss is 14% smaller from Debye screening than from the dynamic screening model. Second, at the low temperatures interestingly both phonon models yield enhanced rather than slowed power dissipation when *dynamic* screening is included. That is, *anti-screening* occurs when the electron temperature T_e is lower than 840 K for the DC model or T_e<770 K for the 3DP model. This is explained as follows. Expressions (13) and (14) show that mathematically the power dissipation is a sum of the contributions that are connected with the various in-plane phonon wavevectors **q**. When dynamic screening is taken into account, the bare electron-phonon interaction is screened or anti-screened depending on wavevector [35]. The large-q components in the summation contribute screening while the small-q components which are connected to the slow motion of the electron gas contribute anti-screening

20

[11], due to the dispersion of the quasi-2D electron plasma frequency. At the low electron temperatures, the small-q components dominate as the degenerate distribution of the dense electrons favours the electron-phonon scattering with small transfer wave-vectors q. At the high temperatures, on the other hand, the electrons are distributed over a large k-space, and they cause screening when the large q-components dominate. We note that antiscreening for the quasi-2D electron energy relaxation was observed early in GaAs quantum wells and occurred also at low electron temperatures [40]. Third, for the three cases with no hot phonons, namely (i), (iii), (iv), the DC model yields higher power dissipation than the 3DP model. However taking into account hot phonons, namely in the corresponding cases (ii), (v), (vi) the 3DP power loss becomes larger. What causes this has become clear after our discussion above for Fig. 3(b); that is, it is due to the higher-energy interface phonons. Fourth, throughout the temperature range the 3DP power loss is 5% smaller than the DC power dissipation in the simplest case (i), but becomes larger (by less than 4%) after including both HPE and screening as in cases (v), (vi). A similar deviation is obtained when reducing the phonon lifetime to 0.1 ps except for the static screening case where only a tiny 0.2% deviation occurs. The 3DP model yields such a close result to the DC calculation, because including screening does not alter the order of the DC and 3DP power loss values in terms of their relative magnitude [that is, the DC power dissipation is higher. Refer to cases (i), (iii), (iv)], while accounting for the HPE does alter the order [compare cases (i), (ii)].

Experimentally, using the microwave noise technique the electron temperature T_e as a function of the supplied power P_s was directly measured for Si-doped bulk GaN [41], strained AlGaN/GaN [9] and lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ [2] heterostructures. The total number of electrons was estimated from the measured low-field Hall mobility and channel resistance using Ohm's law. Under steady-state conditions, the supplied power is equal to the *total* power dissipated to the lattice by the hot electrons. Then one can obtain the experimental data of the *average* power dissipation per electron versus electron temperature (see Fig. 4 of Ref.[2] for the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure). The power loss was shown to increase with the electron temperature (from 2 nW/electron at T_e=500 K, for instance, to 150 nW/electron at T_e=2500 K) but the dependence is complicated by electron screening and the variation of the polar optical phonon lifetime with T_e, as was discussed in our previous study [15]. In the simple approximation where neither hot

phonons nor screening is included, the calculated power dissipation is four times as large as the experimental values. Accounting for screening and HPE brings the calculation much closer to the experimental data. However the theoretical values remain over 2.5 times higher in the low temperature region even with static screening which is generally believed to overestimate the screening effect, and the use of large phonon lifetimes there such as 20 ps will produce a fit with experiment.

To quantify the screening effect and/or HPE a reduction factor β is introduced, $\beta = P_d^0/P_d$, where P_d^0 is the power dissipation without screening or hot phonons, and P_d is the corresponding power loss when screening and/or the hot phonons are included. Fig. 6 shows the temperature-dependences of the reduction factors associated with only the HPE, only screening (Debye screening or dynamic screening), and both HPE and screening calculated with the DC and 3DP phonon models as labeled (using a polar optical phonon lifetime of 1 ps) for the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure. When only screening is included almost the same reduction factors are obtained from the two phonon models, with the two curves coinciding for either static or dynamic screening case. We see antiscreening again in the dynamic screening alone case for electron temperatures lower than about 800 K (below the dotted horizontal line $\beta=1$ in Fig. 6) with reduction factor $\beta < 1$, as antiscreening causes faster power dissipation P_d than P_d^0, $P_d > P_d^0$. In all the other cases, as $\beta > 1$ the electron energy relaxation has slowed down after including screening and/or hot phonons. We see a stronger HPE at low electron temperatures, with the reduction factors decreasing with increasing the electron temperature. The 3DP model underestimates the HPE as expected, and as a result the reduction factors from the 3DP calculation are smaller than the DC result even when screening is included. The reduction factor is $\sim 9\%$ smaller by the 3DP approach than by the DC model in the high-T_e region. Using either of the two screening models the reduction factor associated with both hot phonons and screening decreases as T_e increases, in both the DC and 3DP calculations, but the high-T_e reduction factor tends to be flat and the β values from the static and dynamic screening models are quite close, with β varying only from 2.5 to 3.2. That is to say, with a phonon lifetime of 1 ps the high-temperature electron power loss is reduced approximately by a factor of 3 due to screening and the HPE.

We now turn to the energy relaxation time τ_E. Fig. 7 shows the dependences of the electron energy relaxation times on the electron temperature for the lattice-matched het-

erostructure, calculated with the DC and 3DP models for three cases, namely, (i) excluding hot phonons and screening, (ii) including both the HPE and static (Debye) screening, and (iii) including both HPE and dynamic screening. With no hot phonons or screening, both phonon models yield relaxation times around 0.05 ps but at low electron temperatures $T_e < 500$ K there is a drop in relaxation time τ_E^0 (τ_E^0 is the energy relaxation time with no screening or HPE, $\tau_E^0 = \frac{k_B(T_e-T_0)}{P_d^0}$), which is caused by the exponential rise of the generation number $\Delta N(T_e)$ with T_e. When screening and hot phonons are taken into account, the energy relaxation time can be conveniently expressed as $\tau_E = \frac{k_B(T_e-T_0)}{P_d^0} \frac{P_d^0}{P_d} = \tau_E^0 \beta$, where β is the reduction factor caused by screening and HPE. There is a very small difference in the relaxation times calculated from the two phonon models (upper two pair of curves with $\tau_p=1$ ps in Fig. 7), with the 3DP relaxation times being 4% smaller than the DC ones at high electron temperatures (Fig. 8) and the deviation staying within 5% when reducing the phonon lifetime to 0.1 ps. As the combined hot phonon and screening effect, parametrized by the reduction factor, decreases as T_e is elevated (refer to Fig. 6 above), a great fall of energy relaxation time appears at temperatures $T_e < 750$ K (Fig. 7). In particular the fall is sharp when Debye screening is used as the screening wavenumber κ_D ($\kappa_D \approx 5.8 k_0 T_0/T_e$) decreases fast with T_e (at $T_e=1000$ K, for instance, κ_D reduces to $1.7 k_0$).

At *high* temperatures (above 1200 K), on the other hand, the relaxation time, τ_E, stays almost flat with a very small and slow rise when a single phonon lifetime 1 ps is used throughout the temperature range. The relaxation time is 0.15 ps when static screening is used, which is slightly larger than the $\sim$0.12 ps value obtained with dynamic screening. This saturation in energy relaxation means that the increases in the average electron kinetic energy and power dissipation with T_e are somewhat balanced. Experimentally, saturation in energy relaxation was observed in Si-doped bulk GaN [41] and a strained AlGaN/GaN heterostructure [9] . Experimental results [2, 6] indicate that the high temperature side has phonon life-times τ_p one order of magnitude shorter than 1 ps. Using $\tau_p=0.1$ ps reduces the relaxation time τ_E to $\sim$0.12 ps for the static screening case (thick dashed line in Fig. 7) and to $\sim$0.09 ps when dynamic screening is accounted for (thick dotted line in Fig. 7), as the hot phonon occupation numbers are reduced in the phonon re-absorption processes compared to the case of the longer lifetime of 1 ps. This rapid relaxation means no bottleneck for the power dissipation. Our calculated value $\sim$0.09 ps is nearly equal to the measured high-temperature relaxation time of 0.09 ps [2]. We note that in this case, despite it being

weakened, the hot-phonon reabsorption should be included to obtain the relaxation time $\sim$0.09 ps, as we found that without HPE the relaxation would be faster with $\tau_E \approx$0.06 ps.

At high electron temperatures the measured relaxation time for the investigated lattice-matched heterostructure was found to decrease slowly with the electron temperature [2]. Our calculation shows that the one order of magnitude shorter phonon lifetime has reduced the relaxation time τ_E by only 0.3 ps when static or dynamic screening is accounted for, and therefore the high temperature relaxation time decreases slowly with T_e, in good agreement with experiment. These results also support the experimental finding [6] that the polar optical phonons have a shorter lifetime at a higher electron temperature, otherwise saturation in energy relaxation would occur, similar to that in bulk GaN [7, 41].

We make a comparison of the energy relaxation in bulk GaN [7, 41] and the heterostructure. When hot phonons are ignored, the electron power loss is much greater in bulk GaN, which is approximately three times the power dissipation in the heterostructure when static screening is included. This is largely because the electron density of states is much higher in bulk than in the heterostructure. However we found that the hot phonons play an important role in determining the high-temperature energy relaxation. For Si-doped bulk GaN with a volume electron density 10^{18} cm^{-3}, the high-T_e relaxation time is around 0.2 ps [Fig. 7(a) of Ref.[7]] with phonon lifetime 10 ps, which is longer than the relaxation time of $\sim$0.1 ps in the lattice-matched heterostructure. With a higher electron density, 10^{19} cm^{-3}, for instance, in bulk GaN the electron energy relaxation is found to be much slower due to the combined screening and hot-phonon effect [7]. Therefore, the rapid energy relaxation with τ_E around 0.1 ps in the *heterostructure* means an efficient heat transfer from the hot electron gas to the lattice, which provides the heterostructure with an advantage to use in HFET devices.

V. CONCLUSIONS

In conclusion, we have studied energy relaxation for hot electrons in the quasi-2D channel of lattice-matched InAlN/AlN/GaN heterostructures using the DC and 3DP models. The temperature of the quasi-2D electron gas in the narrow 44-$\mathring{A}$ channel can reach above 2500 K due to high electric power, much higher than the lattice temperature (room temperature). In this study therefore non-equilibrium polar optical phonons as well as electron degeneracy and screening from the mobile electrons are taken into account. Particular attention is

paid to the effects of the two phonon models on the hot-electron relaxation process in the GaN heterostructures. We calculated the electron temperature dependences of the electron power dissipation and energy relaxation time using a variety of phonon lifetime values and examined the 3DP model by comparing the results calculated with the two phonon models. We found that the 3DP model yields very close results to the DC model: with no hot phonons or screening the power loss calculated from the 3DP model is 5% smaller than the DC power dissipation, whereas slightly larger 3DP power loss (by less than 4% with a phonon lifetime from 0.1 to 1 ps) is obtained throughout the electron temperature range after including both the HPE and screening. Very close results are obtained also for the energy relaxation time with the two phonon models (with a percent deviation smaller than 5%). As the investigated heterostructure has a channel narrower than the usual GaN-based heterostructures, therefore the 3DP phonon model is generally a good approximation to use for the study of the energy relaxation in GaN-based heterostructures. We found that our results in the pure phonon emission case are consistent with the sum rules given by Mori and Ando [16] and by Register [17]. The discrepancy between the DC and 3DP results is caused by how much the high energy interface phonons contribute to the energy relaxation: their contribution is enhanced in the pure emission process but is dramatically reduced after including the HPE. Debye screening overestimates the high-T_e energy relaxation time by $\sim$0.03 ps compared to the dynamic screening model whereas with dynamic screening included anti-screening occurs at low electron temperatures (below $\sim$800 K) due to the dispersion of the quasi-2D electron plasma frequency. Our calculation with both phonon models has obtained a great fall in energy relaxation time τ_E at low electron temperatures ($T_e <$ 750 K) and slow decrease at the high temperatures with the use of decreasing phonon lifetime with T_e. The calculated temperature dependence of the relaxation time and the high-temperature relaxation time $\sim$0.09 ps are in good agreement with experimental results. We also compared the quasi-2D hot-electron relaxation with the electron relaxation in bulk GaN and found that the hot phonons play a key role in slowing down the high-T_e electron relaxation for bulk ($\tau_E \sim$ 0.2 ps). For the heterostructures, in contrast, the rapid energy relaxation ($\tau_E \sim$ 0.09 ps) and sub-picosecond phonon decay provide an advantage which benefits electron transport in the HFET devices by efficiently cooling down the extremely hot electrons to increase the electron mobility.

Acknowledgments

Appendix A: Electron power dissipation based on the DC and 3DP models and the sum rules for electron-phonon scattering

As can be seen from Sec. III, $1/\tau_n(q)$ and $1/\tau_{q_z}(q)$ are key quantities in the calculation of the power dissipation. First we look at how to treat $1/\tau_n(q)$ due to interface phonon scattering [Eq. (8)]. Express the matrix element $M_{n,\mathbf{q}}(\mathbf{k};\mathbf{k}+\mathbf{q})$ in terms of the electron and phonon envelope functions, and substitute it into Eq. (8). Then replace $\sum_{\mathbf{k}}$ by $\frac{A}{(2\pi)^2}\int k\,dk\,d\theta$ to convert the summation over electron wavevector to a double integral. The integration over angle θ can be performed analytically, reducing the double integral to an integral over k only. After some algebraic manipulation by changing the variable of integration, we find that the final $1/\tau_n(q)$ can be expressed in a simple form as

$$\frac{1}{\tau_n(q)} = \frac{m^*}{2\pi\hbar^3}\gamma_n^2\frac{1}{q}|\Gamma_I(q)|^2\sqrt{\pi}\left(\frac{k_BT_e}{\varepsilon_q}\right)^{1/2}\left[F_{-\frac{1}{2}}(\xi_{nq}) - F_{-\frac{1}{2}}(\zeta_{nq})\right], \qquad (A1)$$

where $\varepsilon_q = \hbar^2q^2/(2m^*)$, $\Gamma_I(q)$ is the electron-interface-phonon overlap integral

$$\Gamma_I(q) = \int_0^\infty \phi^*(z)e^{-qz}\phi(z)dz, \qquad (A2)$$

and $F_{-\frac{1}{2}}(y)$ is the complete Fermi-Dirac integral of order $-1/2$,

$$F_{-\frac{1}{2}}(y) = \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{x^{-1/2}}{1+e^{x-y}}dx. \qquad (A3)$$

ξ_{nq} and ζ_{nq} are two dimensionless quantities, given by

$$\xi_{nq} = (E_F - \Delta_{nq})/(k_BT_e), \quad \zeta_{nq} = \xi_{nq} - \hbar\omega_n/(k_BT_e), \qquad (A4)$$

where Δ_{nq} has the dimension of energy,

$$\Delta_{nq} = (\varepsilon_q - \hbar\omega_n)^2/(4\varepsilon_q). \qquad (A5)$$

26

Similarly, one can obtain from Eq. (12) the $1/\tau_{q_z}(q)$ expression for half-space phonon scattering,

$$\frac{1}{\tau_{q_z}(q)} = \frac{m^*}{2\pi\hbar^3}\gamma_{LO}^2\frac{1}{L_2}\frac{1}{q^2+q_z^2}|\Gamma_H(q_z)|^2\sqrt{\pi}\left(\frac{k_BT_e}{\varepsilon_q}\right)^{1/2}\left[F_{-\frac{1}{2}}(\xi_q) - F_{-\frac{1}{2}}(\zeta_q)\right], \qquad (A6)$$

where $\Gamma_H(q_z)$ is the electron-half-space-phonon overlap integral

$$\Gamma_H(q_z) = \int_0^\infty \phi^*(z)2\sin(q_zz)\phi(z)dz, \qquad (A7)$$

and ξ_q and ζ_q are given by

$$\xi_q = (E_F - \Delta_q)/(k_BT_e), \quad \zeta_q = \xi_q - \hbar\omega_{LO}/(k_BT_e), \qquad (A8)$$

with the energy Δ_q being defined by

$$\Delta_q = (\varepsilon_q - \hbar\omega_{LO})^2/(4\varepsilon_q). \qquad (A9)$$

For bulk LO phonon scattering, the generation rate $1/\tau(\mathbf{Q})$ is given by

$$\frac{1}{\tau(\mathbf{Q})} = \frac{m^*}{\pi\hbar^3}\gamma_{LO}^2\frac{1}{L_1+L_2}\frac{1}{q^2+q_z^2}|\Gamma_B(q_z)|^2\sqrt{\pi}\left(\frac{k_BT_e}{\varepsilon_q}\right)^{1/2}\left[F_{-\frac{1}{2}}(\xi_q) - F_{-\frac{1}{2}}(\zeta_q)\right], \qquad (A10)$$

where $\Gamma_B(q_z)$ is the electron-LO-phonon overlap integral

$$\Gamma_B(q_z) = \int_0^\infty \phi^*(z)e^{-iq_zz}\phi(z)dz. \qquad (A11)$$

In what follows we confine ourselves to a discusion of the sum rules so we ignore screening and hot phonons. We recall that the calculation of the electron *power dissipation* involves summation over q_z for both half-space and bulk phonon modes [Eq. (13) and (14)]. Inserting the $\Gamma_H(q_z)$ expression Eq. (A7) in Eq. (A6) and then performing summation over q_z for the q_z-dependent factor, we obtain

$$\sum_{q_z>0}\frac{1}{L_2}\frac{1}{q^2+q_z^2}|\Gamma_H(q_z)|^2 = \frac{1}{q}F_H(q), \qquad (A12)$$

where $F_H(q)$ is the form factor for half-space phonons as defined by Mori and Ando [16],

$$F_H(q) = \int_0^\infty dz \int_0^\infty dz'\phi^*(z)\phi(z)\left(e^{-q|z-z'|} - e^{-q|z+z'|}\right)\phi^*(z')\phi(z'). \qquad (A13)$$

Note that in obtaining Eq. (A12) we have used the integration formula

$$\int_0^\infty \frac{\cos ax}{\beta^2+x^2}dx = \frac{\pi}{2\beta}e^{-\beta|a|} \qquad (\beta > 0). \qquad (A14)$$

For the 3D bulk modes [Eqs. (A10) and (A11)], similarly one finds

$$\sum_{q_z} \frac{1}{L_1 + L_2} \frac{1}{q^2 + q_z^2} |\Gamma_B(q_z)|^2 = \frac{1}{2q} F_B(q), \tag{A15}$$

with the form factor given by

$$F_B(q) = \int_0^\infty dz \int_0^\infty dz' \phi^*(z)\phi(z)e^{-q|z-z'|}\phi^*(z')\phi(z'). \tag{A16}$$

For the interface modes [Eqs. (A1) and (A2)], it is readily found that the form factor $F_I(q)$ [16] is simply the square of the overlap integral $\Gamma_I(q)$

$$F_I(q) = \left[\int_0^\infty dz \phi^*(z)e^{-qz}\phi(z)\right]^2. \tag{A17}$$

Using Eqs. (A13), (A16) and (A17) then we restore the form factor sum rule given by Mori and Ando [16]

$$F_B(q) = F_H(q) + F_I(q). \tag{A18}$$

We now proceed to applying this sum rule to the power dissipation as obtained from the DC and 3DP models. We first consider a special case, that is, when the quasi-2D channel in GaN is sufficiently wide that the form factor for the interface phonons $F_I(q)$ can be neglected. This leads to the removal of the contribution to the power dissipation from all the interface modes. Inserting the $1/\tau_{q_z}(q)$ and $1/\tau(\mathbf{Q})$ expressions (A6) and (A10) into the power dissipation equations (13) and (14) respectively, and then using the obtained identities Eqs. (A12) and (A15), we find that the two resulting power dissipation expressions given by the DC and 3DP models are *identical*. In the general case, of course one has $F_I(q) \neq 0$, the two phonon models do not yield the same power dissipation, and the difference is caused entirely by how much the true interface modes (eigenmodes) contribute compared to that when they are treated simply as bulk LO modes in terms of the phonon frequency and electron-phonon coupling strength. This can be made more clear by the following proof. If we ignore the difference in the material parameters, such as the phonon frequencies and dielectric constants, of the two constituents of the heterostructure, and use only the GaN parameters, the interface mode frequency that is given by the solution of $\epsilon(\omega) + \bar{\epsilon}(\omega) = 0$ simply reduces to the LO phonon frequency ω_{LO} of GaN, and $\frac{1}{\epsilon_n}$ for the interface modes reduces to $\frac{1}{\epsilon_{LO}}$ for the bulk LO modes, thus making the coupling constant α_n for the interface modes, $\alpha_n = \frac{e^2}{\hbar}(\frac{m^*}{2\hbar\omega_n})^{1/2}/\epsilon_n$ [16], reduce to the Fröhlich coupling constant α_{LO} for the bulk

LO modes, $\alpha_{LO} = \frac{e^2}{\hbar}(\frac{m^*}{2\hbar\omega_{LO}})^{1/2}/\epsilon_{LO}$. This is exactly the scenario that results from the degenerate modes approximation [16, 17]. It then follows that the use of the Mori-Ando sum rule [Eq. (A18)] results in the same DC and 3DP power dissipation. Furthermore, the same electron-phonon scattering rate is also obtained with the two phonon models, which is consistent with the sum rule for the electron-phonon interaction given by Register [17]. Of course the two constituent materials are different, and the true interface modes are not bulk GaN LO modes, and therefore in general the DC power dissipation is different from that evaluated with the 3DP model, even in the non-HPE case, as our numerical results in Sec. IV have shown.

Appendix B: Phonon wavevector dependences of energy Δ_q and the Fermi-Dirac integral $F_{-\frac{1}{2}}(\xi_q)$

The following illustration is made for the half-space modes but it can be equally applied to the interface and bulk LO modes as well. The energy expression appearing in the delta function of Eq. (12) indicates energy conservation when an electron at state $\mathbf{k}$ makes a transition up to state $\mathbf{k} + \mathbf{q}$ after a half-space phonon with wavevector $\mathbf{q}$ is absorbed. If θ is the angle between $\mathbf{k}$ and $\mathbf{q}$, one finds $\cos\theta = \frac{\hbar\omega_{LO} - \varepsilon_q}{\hbar^2 kq/m^*}$. As $|\cos\theta| \leq 1$, one obtains $E_k \geq \Delta_q$ [Δ_q is given by Eq. (A9)]. Therefore, the energy Δ_q, which arises from energy conservation plus momentum conservation, has a clear physical meaning; that is, given a phonon wavevector, Δ_q is the *minimum* kinetic energy of the electron for the phonon absorption to occur. Fig. 9(a) shows the energy Δ_q as a function of the wavevector q, where for simplicity Δ_q and q are made dimensionless with respect to the LO phonon energy $\hbar\omega_{LO}$ and electron wavevector for threshold LO phonon emission k_0, respectively. We see that the energy Δ_q decreases steeply to the minimum value of 0 at $q = k_0$, and then slowly increases as wavevector q becomes larger. Eq. (A6) shows that the generation rate $1/\tau_{q_z}(q)$ is proportional to the difference of the Fermi-Dirac integrals at the two arguments differing by only $\hbar\omega_{LO}/(k_B T_e)$ [also refer to Eq. (A8)], owing to scattering by half-space phonons of energy $\hbar\omega_{LO}$. Further, the Fermi-Dirac integral $F_{-\frac{1}{2}}(y)$ [Eq. (A3)] is a monotonically increasing function, and this ensures that $F_{-\frac{1}{2}}(\xi_q) - F_{-\frac{1}{2}}(\zeta_q)$ and consequently the rate $1/\tau_{q_z}(q)$ is always positive as required physically. Therefore, knowing the dependence of $F_{-\frac{1}{2}}(\xi_q)$ (or $F_{-\frac{1}{2}}(\zeta_q)$ equally) on the wavevector q is fundamental to understanding the q-

dependence of the non-equilibrium phonon occupation number in Sec. IV. As shown in Fig. 9(b) (solid line), $F_{-\frac{1}{2}}(\xi_q)$ increases rapidly from 0 to the maximum value at $q = k_0$, and then decreases with further increasing q. In both limits of $q \to 0$ and $q \to \infty$ the Fermi-Dirac integral is zero leading to null phonon generation. In the non-degenerate Boltzmann approximation, one has $F_{-\frac{1}{2}}(y) = e^y$, making $F_{-\frac{1}{2}}(\xi_q) = e^{\xi_q}$ simply proportional to $e^{-\Delta_q/k_B T_e}$. Then one finds that $F_{-\frac{1}{2}}(\xi_q)$ depends on q according to $e^{-\alpha(\frac{k_0}{q})^2/4}$ for $q \to 0$, and according to $e^{-\alpha(\frac{q}{k_0})^2/4}$ for $q \to \infty$ where $\alpha = \hbar\omega_{LO}/(k_B T_e)$. For an electron density of $7 \times 10^{11}/\text{cm}^2$ at the electron temperature $T_e = 2\hbar\omega_{LO}/k_B$=2110 K, for instance, the non-degeneracy approximation e^{ξ_q} and the Fermi-Dirac integral $F_{-\frac{1}{2}}(\xi_q)$ are very close, making the lower two curves in Fig. 9(b) coincide. In typical GaN based heterostructures the electron density is quite high. For electron density $1.2 \times 10^{13}/\text{cm}^2$, the non-degeneracy approximation (dot-dashed line) is quite large with a peak value of 60% larger than the maximum value of $F_{-\frac{1}{2}}(\xi_q)$ (solid line). Nevertheless the dependence of $F_{-\frac{1}{2}}(\xi_q)$ on the phonon wavevector q governs the q-dependence of the non-equilibrium phonon occupation number, as shown in Sec. IV.

---

[1] H. Morkoç, *Nitride Semiconductors and Devices* (Springer, Heidelberg, 1999).

[2] A. Matulionis, J. Liberis, E. Šermukšnis, J. Xie, J. H. Leach, M. Wu, and H. Morkoç, Semicond. Sci. Technol. **23**, 075048 (2008).

[3] A. Matulionis, J. Liberis, I. Matulionienė, M. Ramonas, E. Šermukšnis, J. H. Leach, M. Wu, X. Ni, X. Li, and H. Morkoç, Appl. Phys. Lett. **95**, 192102 (2009).

[4] J. H. Leach, C. Y. Zhu, M. Wu, X. Ni, X. Li, J. Xie, Ü. Özgür, H. Morkoç, J. Liberis, E. Šermukšnis, et al., Appl. Phys. Lett. **95**, 223504 (2009).

[5] J. H. Leach, C. Y. Zhu, M. Wu, X. Ni, X. Li, J. Xie, Ü. Özgür, H. Morkoç, J. Liberis, E. Šermukšnis, et al., Appl. Phys. Lett. **96**, 133505 (2010).

[6] J. Liberis, I. Matulionienė, A. Matulionis, E. Šermukšnis, J. Xie, J. H. Leach, and H. Morkoç, Phys. Status Solidi A **206**, 1385 (2009).

[7] J.-Z. Zhang, J. Appl. Phys. **115**, 203704 (2014).

[8] J. Khurgin, Y. J. Ding, and D. Jena, Appl. Phys. Lett. **91**, 252104 (2007).

[9] A. Matulionis, J. Liberis, M. Ramonas, I. Matulionienė, L. F. Eastman, A. Vertiatchikh,

X. Chen, and Y.-J. Sun, Phys. Status Solidi C **2**, 2585 (2005).

[10] J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Berlin, 1996).

[11] B. K. Ridley, *Electrons and Phonons in Semiconductor Multilayers* (Cambridge University Press, Cambridge, 2009).

[12] D. R. Anderson, N. A. Zakhleniuk, M. Babiker, B. K. Ridley, and C. R. Bennett, Phys. Rev. B **63**, 245313 (2001).

[13] A. Dyson and B. K. Ridley, J. Appl. Phys. **109**, 054509 (2011).

[14] J.-Z. Zhang, A. Dyson, and B. K. Ridley, Phys. Rev. B **84**, 155310 (2011).

[15] J.-Z. Zhang, A. Dyson, and B. K. Ridley, Appl. Phys. Lett. **102**, 062104 (2013).

[16] N. Mori and T. Ando, Phys. Rev. B **40**, 6175 (1989).

[17] L. F. Register, Phys. Rev. B **45**, 8756 (1992).

[18] S. D. Sarma and B. A. Mason, Ann. Phys. **163**, 78 (1985).

[19] H. Rücker, E. Molinari, and P. Lugli, Phys. Rev. B **45**, 6747 (1992).

[20] J.-Z. Zhang, B.-F. Zhu, and K. Huang, Phys. Rev. B **59**, 13184 (1999).

[21] K. J. Nash, Phys. Rev. B **46**, 7723 (1992).

[22] S. Das Sarma, J. K. Jain, and R. Jalabert, Phys. Rev. B **41**, 3561 (1990).

[23] B. C. Lee, K. W. Kim, M. Dutta, and M. A. Stroscio, Phys. Rev. B **56**, 997 (1997).

[24] A. Dyson and B. K. Ridley, Appl. Phys. Lett. **102**, 042108 (2013).

[25] A. Dyson and B. K. Ridley, J. Appl. Phys. **112**, 063707 (2012).

[26] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, London, 1961).

[27] D. Lowe and J. R. Barker, J. Phys. C: Solid State Phys. **18**, 2507 (1985).

[28] F. F. Fang and W. E. Howard, Phys. Rev. Lett. **16**, 797 (1966).

[29] F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967).

[30] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[31] D. R. Anderson, M. Babiker, C. R. Bennett, N. A. Zakhleniuk, and B. K. Ridley, J. Phys.: Condens. Matter **13**, 5999 (2001).

[32] J. C. Nipko and C.-K. Loong, Phys. Rev. B **57**, 10550 (1998).

[33] C. Bulutay, B. K. Ridley, and N. A. Zakhleniuk, Phys. Rev. B **62**, 15754 (2000).

[34] C. Bulutay, B. K. Ridley, and N. A. Zakhleniuk, Physica B **314**, 63 (2002).

[35] B. K. Ridley, *Quantum Processes in Semiconductors* (Clarendon Press, Oxford, 1999).

[36] B. K. Ridley, J. Phys. C: Solid State Phys. **15**, 5899 (1982).

[37] F. A. Riddoch and B. K. Ridley, J. Phys. C: Solid State Phys. **16**, 6971 (1983).

[38] M. A. Stroscio and M. Dutta, *Phonons in Nanostructures* (Cambridge University Press, Cambridge, 2001).

[39] O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, et al., J. Phys.: Condens. Matter **14**, 3399 (2002).

[40] S. Das Sarma, J. K. Jain, and R. Jalabert, Phys. Rev. B **37**, 6290 (1988).

[41] 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).

FIG. 1: (Color online) (a) Non-equilibrium phonon occupation numbers vs in-plane phonon wavevector of the half-space phonons for a number of q_z indices as labeled ($q_z = l\pi/L_2$, $l \geq 1$, L_2 being the dimension of the GaN half-space), generated from energy relaxation of the hot electrons at temperature 1000 K in a typical lattice-matched heterostructure $In_{0.18}Al_{0.82}N/AlN/GaN$ with an areal electron density of $1.2 \times 10^{13}/cm^2$ (corresponding to an effective channel width of 44 Å), calculated with a polar optical phonon life-time τ_p=0.1 ps. The horizontal dotted line shows the thermal equilibrium occupation number at room temperature, $N(\omega_{LO}, T_0) = 0.03$. (b) Phonon-wavevector-dependent factor, $\Gamma_H^2(q_z)k_0^2/(q^2+q_z^2)$, of the squared electron-phonon interaction matrix elements of half-space modes, as functions of q_z for three in-plane wavevectors q=0.5, 1, 2k_0, where k_0 is the characteristic electron wave-vector for threshold LO phonon emission, $k_0 = \sqrt{2m^*\omega_{LO}/\hbar}$.

FIG. 2: (Color online) Non-equilibrium phonon occupation numbers of the lower-energy (IF_1) and higher-energy (IF_2) interface phonons from energy relaxation of the hot electrons at temperature 1000 K in the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure, calculated with a phonon life-time τ_p=0.1 ps for (a) excluding and (b) including screening with the static (Debye) and dynamic screening models.

FIG. 3: (Color online) Average power dissipated per electron vs electron temperature for (a) a strained $Al_{0.05}Ga_{0.95}N/GaN$ heterostructure with an effective GaN-channel width of 110 Å (corresponding to an areal electron density of $7 \times 10^{11}/cm^2$) and (b) the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure with an effective GaN-channel width of 44 Å (corresponding to the areal electron density of $1.2 \times 10^{13}/cm^2$), calculated with the dielectric continuum (DC) and three-dimensional phonon (3DP) models for excluding or including the hot-phonon effect (HPE) with two phonon life-time values τ_p=0.5 and 2 ps as labeled.

FIG. 4: (Color online) (a) Average power dissipated per electron vs electron temperature in the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure, due to scattering with the half-space (HS), lower-energy (IF_1) and higher-energy (IF_2) interface phonons, respectively, and (b) higher-energy interface phonon (IF_2) generation number per unit time $W_2(q)$ [Eq. (7)] vs in-plane phonon wave-vector at the electron temperature 2500 K, when the hot-phonon effect (HPE) is excluded or included with a phonon life-time of τ_p=0.5 ps. In (b) the phonon generation $W_2(q)$ values of including the hot phonons have been enlarged by ten times.

FIG. 5: (Color online) Average power dissipated per electron vs electron temperature in the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure, calculated with the dielectric continuum (DC) and three-dimensional phonon (3DP) models for the six cases as labeled, namely, (i) without hot-phonon effect (HPE) or screening (scr), (ii) including only the HPE, (iii) including only static (Debye) screening, (iv) including only dynamic screening, (v) including both the HPE and static screening, and (vi) including both the HPE and dynamic screening. The polar optical phonon life-time of 1 ps is used when the hot phonons are included for the three cases (ii), (v), (vi).

FIG. 6: (Color online) Reduction factors versus electron temperature associated with only the hot-phonon effect (HPE), only screening (Debye screening or dynamic screening), and both HPE and screening calculated with the dielectric continuum (DC) and three-dimensional phonon (3DP) models as labeled, where a polar optical phonon lifetime of 1 ps is used for hot phonons in the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure. Also drawn is a dotted horizontal line β=1, the part of curves below which indicates antiscreening.

FIG. 7: (Color online) Energy relaxation times vs electron temperature of the hot electrons in the lattice-matched $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure, calculated with the dielectric continuum (DC) and three-dimensional phonon (3DP) models for several cases as labeled: with no hot-phonon effect (HPE) or screening (scr), including both HPE and static (Debye) screening, and including both the HPE and dynamic screening with a phonon lifetime of $\tau_p=1$ ps. The DC results at high electron temperatures using both screening models calculated with the phonon lifetime 0.1 ps are also shown.

FIG. 8: (Color online) Percent deviations of the energy relaxation times evaluated with the three-dimensional phonon (3DP) approximation with respect to the dielectric continuum (DC) calculation as functions of the electron temperature in the $In_{0.18}Al_{0.82}N/AlN/GaN$ heterostructure for several cases as indicated: with no hot-phonon effect (HPE) or screening (scr), including both the HPE and static (Debye) screening, and including both the HPE and dynamic screening with a phonon lifetime of $\tau_p=1$ ps. The relative deviations at high electron temperatures using both screening models calculated with the phonon lifetime 0.1 ps are also shown.

FIG. 9: (Color online) (a) The energy Δ_q, namely, the *minimum* electron kinetic energy for phonon absorption, given by $\frac{\Delta_q}{\hbar\omega_{LO}} = \frac{[(\frac{q}{k_0})^2-1]^2}{4(\frac{q}{k_0})^2}$, as a function of the phonon wavevector q; (b) the Fermi-Dirac integral $F_{-\frac{1}{2}}(\xi_q)$, where $\xi_q = (E_F - \Delta_q)/(k_B T_e)$, as functions of the phonon wavevector q for the two electron densities 1.2×10^{13} (solid) and 7×10^{11} cm^{-2} (dotted) at the same electron temperature $T_e = 2\hbar\omega_{LO}/k_B$=2110 K. Also shown is the non-degenerate Boltzmann approximation, e^{ξ_q}, for the Fermi-Dirac integral $F_{-\frac{1}{2}}(\xi_q)$ for the two electron densities respectively (dot-dashed and dashed). Note that k_0 is the characteristic electron wave-vector for threshold LO phonon emission, $k_0 = \sqrt{2m^*\omega_{LO}/\hbar}$.
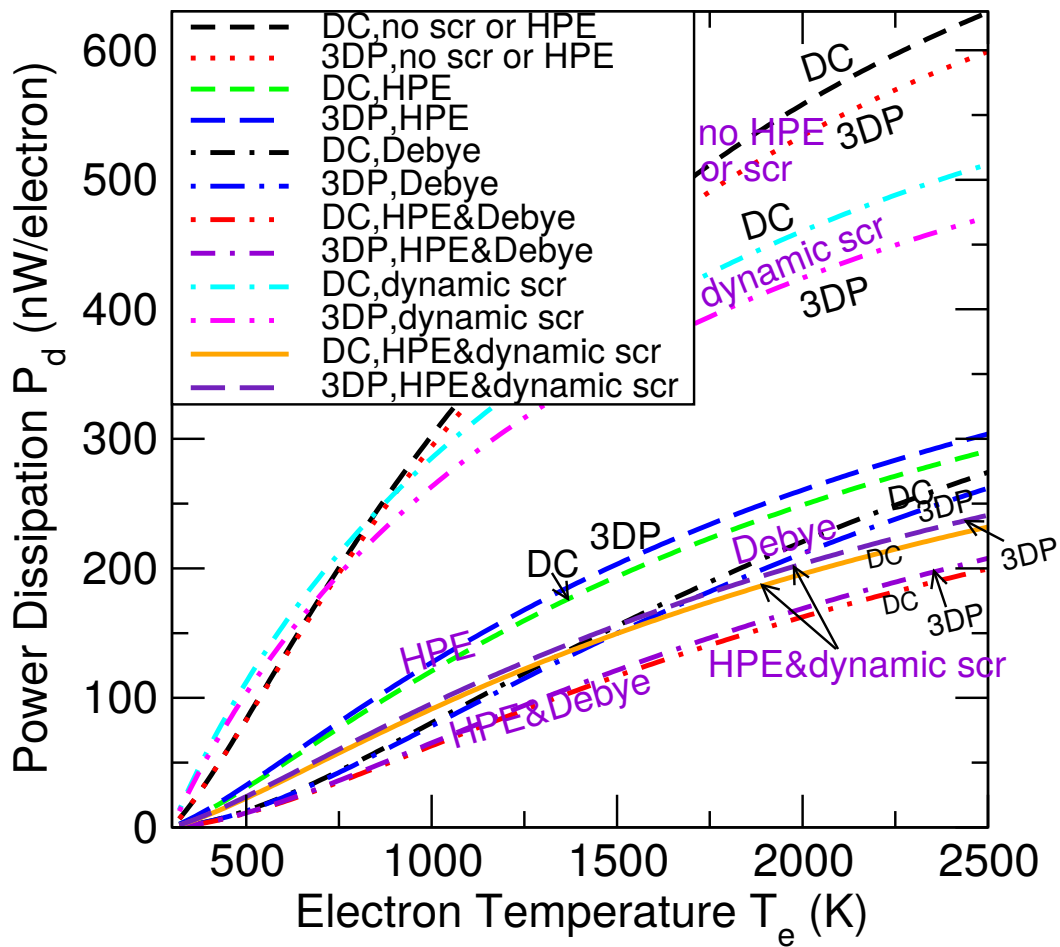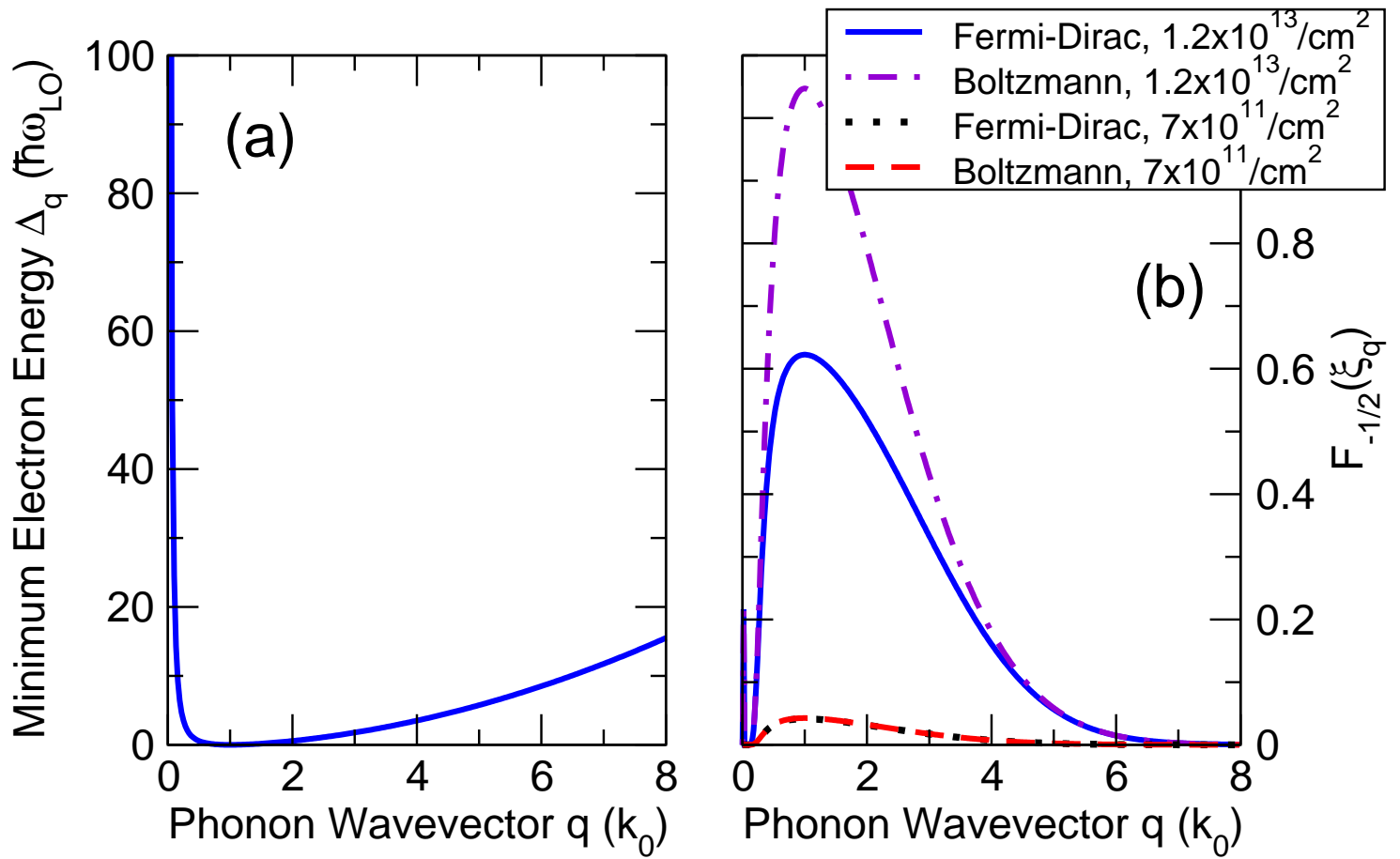
Fig. 5

Fig. 9