Polarization-mediated remote surface roughness scattering in ultrathin barrier GaN high-electron mobility transistors

Yu Cao, Huili Xing, and Debdeep Jena
Department of Electrical Engineering, University of Notre Dame, Indiana 46556, USA

(Received 26 October 2010; accepted 5 November 2010; published online 2 December 2010)

Ultrathin AlN/GaN heterojunctions are highly attractive for high-frequency transistor applications. In this work, remote surface roughness (RSR) scattering mediated by the high polarization is studied as a new scattering mechanism in such structures. In both depletion-mode and enhancement-mode devices with ultrathin AlN barriers, RSR scattering can be of the same order as polar optical phonons, the dominant scattering mechanism at room temperature. The study indicates that to achieve high-performance high-electron mobility transistors, the surface roughness during processing is as critical as maintaining a sharp heterojunction. © 2010 American Institute of Physics. [doi:10.1063/1.3521258]

III-V nitride-based high-electron mobility transistor (HEMT) technology has made rapid progress over the last decade. For future high-speed device applications, both lateral and vertical scalings are necessary. In traditional AlGaN/GaN HEMTs, when the AlGaN barrier thickness is reduced, the 2DEG density decreases, and the access resistance increases, which degrades high-frequency performance. AlN/GaN heterojunctions provide extremely high-electron densities, which degrades high-frequency performance. In traditional AlGaN/GaN HEMTs, when the AlGaN barrier thickness is reduced, the 2DEG density decreases, and the access resistance increases, which degrades high-frequency performance. AlN/GaN heterojunctions provide extremely high-electron densities at the heterointerface covered by only 3–4 nm AlN barrier,1 making them attractive for high-frequency applications.

Transport properties of the two dimensional electron gases (2DEGs) at such heterojunctions have been studied extensively to understand the performance limits of nitride devices. In this paper, a new scattering mechanism caused by remote surface roughness (RSR) is discussed. A similar concept has been studied for silicon metal-oxide-semiconductor field-effect transistors (MOSFETs),2,3 where the variation of the oxide thickness causes the degradation of electron mobility in the channel under gate bias. In AlN/GaN HEMTs, AlN is an epitaxial gate dielectric. RSR scattering originates from the polarization in the barrier material and exists even when no gate bias is applied. The origin of such scattering is due to a rough surface coupled with polarization. Surface roughness is common in epitaxy; this scattering mechanism is valid for 2DEGs generated in all polar semiconductor heterostructures including II-VI oxide ZnO/MgZnO (Ref. 4) and especially perovskite oxides heterostructures5,6 with very high polarization. Al(Ga)N/GaN HEMTs are a specific case treated in this work. Owing to the very large bandgap, AlN barriers allow a high degree of vertical scaling than silicon MOSFETs that incorporate high-K dielectrics; thus this scattering mechanism is especially severe for ultrathin AlN barrier HEMTs.

In polar semiconductor heterostructures, the 2DEG density is a function of the barrier thickness. A Schottky contact with a gate metal makes the surface of the barrier an equipotential plane. However, thickness variations in the barrier result in variations of the electric potential in the channel plane; this random potential variation is the origin of the scattering mechanism. To understand the effect of RSR scattering on device performance, it is necessary to study RSR scattering in the channel region under the gate. In depletion-mode (D-mode) HEMTs, the AlN barrier thickness exceeds the minimum thickness for 2DEG formation at zero gate bias. Due to remote surface roughness, the 2DEG density is not uniform along the channel. This will lead to different threshold voltages $V_{th}$ across large wafers. Enhancement-mode (E-mode) HEMTs are desired for high power switches and high temperature integrated circuits. Gate recessing is one of the most commonly used approaches to fabricate E-mode HEMTs.7,8 An ultrathin AlN interlayer is widely used in AlGaN or InAlN HEMTs to remove alloy scattering. With gate recessing, dry etching techniques have been developed to selectively remove the top barrier, such as InAlN, and the gate metal is directly deposited on the ultrathin AlN interlayer.9 The RSR scattering model derived in this paper is especially important in such E-mode HEMTs with recessed AlGaN or InAlN barriers.

Since the scattering originates from the thickness variation of the AlN barrier, it is necessary to find the functional dependence of the 2DEG density $n_z$ in the gated channel on the barrier thickness $t$ and the gate bias $V_g$. To do so, from the energy band diagram of an AlN/GaN heterojunction (Fig. 1), the electric field in the barrier can be expressed as

$$F_b = eV_b$$

$$eV_g$$

$$E_F$$

$$\Delta E_C$$

$$0$$

$2DEG$ $\text{AlN}$ $\text{GaN}$

FIG. 1. (Color online) Band diagram of the AlN/GaN heterojunction calculated by 1D Poisson–Schrödinger solver.
$$F_b = -e(n_s - n_g) = -rac{e(2E_b - V_g) + E_F - E_c}{e},$$

(1)

where $e$ is the electron charge, $n_s$ is the 2DEG density in the well, $V_g$ is the surface barrier height, $E_c$ is the conduction band offset between the barrier and the well, $e$ is the dielectric constant of the barrier material, $t$ is the thickness, and $E_F$ is the Fermi energy shown in Fig. 1. The electron wave function in the growth direction $z$ is well approximated by the Fang–Howard variational form,

$$\phi(z) = \phi_0(z) = \phi_{10}(z) = \phi_{20}^2 \exp(-b z^2),$$

where $b$ is given by $b = (33m^* e^2 n_s / 8h^2 \epsilon_0 \epsilon_{A})^{1 \over 3}$. The ground-state energy $E_0$ is

$$E_0 = \left[ \frac{5}{16} \left( \frac{33}{2} \right)^{2/3} \right] \frac{h^2}{2 m^*} \left( \frac{e^2 n_s}{2 \epsilon_0 \epsilon_{AI}^b} \right)^{2 / 3},$$

(2)

where $F_w$ is the electric field in the well, $\epsilon_w$ is the dielectric constant of the well material, and $m^*$ is the effective mass in the well. If the electron density in the second sub-band is negligible compared to that in the first sub-band (which is the case in AlN/GaN heterojunctions with very thin AlN barriers), the electron density can be approximated by the electric quantum limit

$$n_s = \frac{m^*}{2 \pi \hbar^2} (E_F - E_0).$$

(3)

Multiple subband occupation should be considered at high gate overdrives when the channel electron density is large, but in this work, we restrict the treatment to the electric quantum limit. This is valid when the electric density in the channel is low and the channel is close to pinch-off; indeed the speed and transconductance of GaN HEMTs peak near this bias point. Combining Eqs. (2) and (3), we can write $E_F$ as a function of $n_s$

$$E_F = \frac{\pi \hbar^2}{m^*} n_s + 2.025 \left[ \frac{h^2}{2 m^*} \left( \frac{e^2 n_s}{2 \epsilon_0 \epsilon_{AI}^b} \right)^{2 / 3} \right].$$

(4)

Substituting Eq. (4) into Eq. (1), we obtain the required relation between $n_s$, $t$, and $V_g$,

$$\frac{e^2 (\sigma - n_g)}{\epsilon_0 \epsilon_{AI}^b} = \frac{e(V_b - V_g) - E_c + \pi \hbar^2 n_s}{m^*} + 2.025 \left[ \frac{h^2}{2 m^*} \left( \frac{e^2 n_s}{2 \epsilon_0 \epsilon_{AI}^b} \right)^{2 / 3} \right].$$

The equation is cubic in $n_s$. For a given $V_g$, the 2DEG density $n_s$ can be either analytically or numerically solved as a function of the barrier thickness $t$. Figure 2(a) shows that the calculated 2DEG density $n_s$ as a function of the gate bias $V_g$. A fixed surface barrier height $eV_b = 3$ eV (Ref. 1) and AlN barrier thicknesses of $t=2$, 3, and 4 nm are used for the calculation. $t=1$ nm is used for the calculation in E-mode devices later. The different slopes in Fig. 2(a) are indicative of the different capacitances.

The electron volume density along the $z$ direction is

$$\rho(z) = -en_s \bar{\rho} = -en_s (b^2/2) \exp(-b z^2).$$

Identifying the heterointerface as $z=0$, the electric field $F_w(z)$ in the well $\rightarrow 0$ when $z \rightarrow \infty$. The electric field variation $F_w(z)$ in the triangular quantum well is thus

$$F_w(z) = \int F(z) \, dz = \int \frac{e n_s \bar{\rho}}{\epsilon_0 \epsilon_{AI}^b} \, dz = \frac{e n_s \bar{\rho}}{\epsilon_0 \epsilon_{AI}^b} \int_0^\infty \frac{e n_s \bar{\rho}}{\epsilon_0 \epsilon_{AI}^b} \, dz = \frac{e n_s \bar{\rho}}{\epsilon_0 \epsilon_{AI}^b} \int_0^\infty \exp(-b z^2) \, dz = \frac{\sqrt{\pi} n_s \bar{\rho}}{\sqrt{2} \epsilon_0 \epsilon_{AI}^b} = \frac{\sqrt{\pi} \rho_{\text{avg}}}{8 \epsilon_0 \epsilon_{AI}^b},$$

where $\rho(z)$ is the electron density along the $z$ direction, $\rho_{\text{avg}}$ is the average electron density, $\Omega$ is the area in $x$-$y$ plane, and $F_{\text{avg}} = P_0 + P_1/8 + (3 P_2/16 b) + (15 P_3/16 b^3)$. The total scattering rate due to remote roughness scattering is given by
FIG. 3. (Color online) In E-mode devices, assuming that Gaussian correlation is applied with screening effect considered, mobilities given by RSR scattering are plotted (a) against different correlation lengths $L$ under the gate bias $V_g$=0.75 V (solid lines) and $V_g$=2 V (dashed lines) with $\Delta$ =0.5 nm (2 MLs) and $\Delta$=0.25 nm (1 ML). Mobilities limited by POP scattering at RT with corresponding $V_g$ are also plotted. (b) Mobilities given by RSR scattering are plotted against the gate bias $V_g$ with $L$=2 nm, $\Delta$=0.5 nm, and $t$=1 nm, where the mobilities limited by POP and AP scattering at RT are also plotted, as well as the total mobility calculated with Matheissen’s rule applied.

$$\frac{1}{\tau_{RSR}} = \frac{\pi e^2 m^*}{\hbar^2} \Delta L^2 F^2_{Z} \text{avg} e^{-Z} [I_g(Z) - I_t(Z)],$$

where $Z=\pi n L^2$, $I_g(Z)$, and $I_t(Z)$ are modified Bessel functions of the first kind and are the solutions of $Z^2 (d^2/\text{d}z^2) +2 (d/\text{d}z) - (Z^2 + \nu^2) f = 0$, with $\nu$ equal to 0 and 1 respectively. We note that the RSR scattering rate increases as the square of the roughness $\Delta$, the correlation length $L$, and the polarization charge $\sigma_p$.

Figure 2(b) shows the calculated electron mobilities limited by RSR scattering in a D-mode AlN HEMT with a barrier thickness $t_{AlN} \sim 2$ nm. At a negative gate bias of $V_g$=−0.5 V, the 2DEG density in the channel is $n_\parallel \sim 9 \times 10^{12}$ cm$^{-2}$. The RSR scattering rate is found to be of the same order of magnitude (2−3×) as polar optical phonon (POP) scattering, which is the strongest RT scattering mechanism. This result indicates that under negative gate bias, RSR scattering can degrade the 2DEG mobilities in D-Mode HEMTs with thin AlN barriers. Thus, either a smoother surface or a thicker barrier is preferred in D-mode devices for the uniformity of 2DEG transport properties and high mobilities.

In recently reported E-mode InAlN HEMTs, the AlN barrier is only $t=1$ nm thick under the gate. With such barrier, mobilities limited by RSR scattering under different gate biases $V_g$ (0.75 V in solid lines and 2.0 V in dashed lines) are plotted in Fig. 3(a), together with mobilities limited by POP scattering. With an increasing gate overdrive, the channel charge density increases, leading to a smaller electric field in the AlN barrier. Thus, the potential fluctuation due to the remote surface roughness is less for the same $\Delta$. Therefore, the 2DEG experiences less RSR scattering at high $V_g$. POP scattering rate increases with higher $V_g$ due to the increased $n_\parallel$. To study the effect of the surface roughness $\Delta$ on RSR scattering, $\Delta$=0.25 nm (1 ML) and 0.5 nm (2 MLs) are selected. When the surface is rough ($\Delta$=0.5 nm, $V_g$ =0.75 V, and $L$=2 nm), the mobility is ~2000 cm$^2$/V s limited by POP scattering and ~4000 cm$^2$/V s limited by RSR scattering. The total mobility can be degraded from 2000 to ~1300 cm$^2$/V s due to the presence of RSR scattering. If smaller roughness is achieved, e.g., $\Delta$=0.25 nm (~1 ML), RSR scattering can be suppressed and POP scattering is still the major scattering mechanism that limits 2DEGs mobility in the channel. This indicates that for E-mode HEMTs using gate recessing technique, a low- damage etching process is critical for achieving high channel mobilities.

Figure 3(b) shows the $V_g$ dependence of the mobility in E-mode HEMTs with $t$=1 nm, $\Delta$=2 MLs, and $L$=2 nm. RSR scattering is more severe than acoustic phonon (AP) scattering. At $V_g$=0.7 V, the mobility limited by RSR scattering is close to that limited by POP scattering, which is the dominant scattering mechanism in nitride HEMTs at RT. Due to the presence of RSR scattering, the total mobility reduces from ~2000 cm$^2$/V s given by POP scattering to ~1000 cm$^2$/V s, indicating a substantial degradation.

In summary, the combined effect of high polarization and ultrathin barriers can lead to strong degradation of mobility in 2DEGs in the presence of surface roughness. In highly polar heterostructures with thin barriers, this scattering is severe and comparable to optical phonon scattering at RT. The study of RSR scattering in ultrascaled E-mode GaN HEMTs indicates that a uniform barrier thickness is highly desired to avoid the degradation of 2DEG mobilities. The model also applies to 2DEG transport study in other polar heterostructures. To minimize the influence of RSR scattering in E-mode HEMTs, the electric field in the barrier $F_b$ has to be reduced. A thin layer of dielectric between gate metal and recessed AlN barrier can effectively reduce the $F_b$ as well as gate leakage but with a penalty of a lower transconductance. Since the highest transconductance $g_m$ and cut-off frequency $f_c$ of GaN HEMTs occurs near pinch-off voltage, RSR scattering will assume an increasing important role in future design of high-speed GaN HEMTs. Improved epitaxy and processing techniques that preserve a smooth surface and a uniform barrier thickness will be critical.