Mechanisms of gate lag in GaN/AlGaN/GaN high electron mobility transistors

Oleg Mitrofanov*, Michael Manfra

Bell Laboratories, Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974, USA

Received 11 December 2003; received in revised form 22 December 2003; accepted 22 December 2003

Available online 21 February 2004

Abstract

The presence of electronic traps in GaN-based transistors limits device performance and reliability. It is believed that material defects and electronic states on GaN surface act as the trapping centers. In spite of extensive investigation of trapping phenomena, the physics of the active defects is not completely understood. Charge trapping in the device structure is reflected in gate lag, a delayed response of the channel current to modulation of the gate potential. Gate lag studies provide essential information about the traps allowing identification of the active defects. In this paper we review gate lag in GaN-based high electron mobility transistors (HEMTs). Current transient spectroscopy, a characterization method based on gate lag measurements, is applied for trap identification in AlGaN/GaN HEMTs grown by plasma-assisted molecular beam epitaxy. In particular we focus on the processes of electron capture and emission from the traps. Probing the charge transfer mechanisms leading to gate lag allows us to extract the trap characteristics including the trapping potential, the binding energy of an electron on the trap, and the physical location of the active centers in the device.

© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

GaN-based electronic devices have recently demonstrated excellent performance at microwave frequencies. State of the art AlGaN/GaN high electron mobility transistors (HEMTs) were shown to produce up to 12 W mm$^{-1}$ at 2 GHz [1]. While the achieved power density demonstrates tremendous potential for GaN in a variety of applications, the current state of GaN RF power devices has not yet matured to the point where...
they have replaced existing technologies. One of the major issues that continues to limit the performance of GaN-based devices is the presence of electronic traps in the device structure. In AlGaN/GaN HEMTs, the parasitic charge moving in and out of the traps on the surface and/or in the bulk of the heterostructure affects the density of the two dimensional electron gas (2DEG) in the channel, causing effects such as current collapse [2, 3], and transconductance frequency dispersion [4–6]. The characteristic time of the recharging process in GaN ranges between nanoseconds and seconds. As a result, the trapping effects can limit device performance even at relatively low frequencies. In addition, the thermally activated traps contribute significantly to the device low-frequency noise [7, 8].

Understanding the origin of the traps in GaN-based transistors, their location, and the physical mechanisms involved in the trapping is important for the optimization of device performance. Currently, the trapping processes in GaN are not completely understood, in spite of considerable research effort that has been directed toward identification and elimination of the traps [9–21]. The majority of these studies provide only qualitative and often contradicting explanations of the trapping phenomena. This inconsistency existing in the field is largely related to the diversity of the trapping effects in GaN and varying material quality. GaN contains high densities of defects and dislocations formed during the growth due to the large difference in lattice constants and in thermal expansion coefficients of the substrate and the epilayers. The defects and dislocations can potentially act as the charge carrier traps creating localized levels inside the bandgap. In addition, it is believed that the surface of the material contains a large density (>10^{13} \text{ cm}^{-2}) of donor-like states [22]. While the majority of the trapping effects result in similar degradation of the transistor characteristics at high frequencies, the dominating trapping mechanisms could vary in devices grown by different methods or subjected to different processing procedures. It is essential, therefore, that any characterization method differentiate between various trapping centers.

Transient spectroscopy allows extraction of the activation energy of the trap and the trap cross-section [23]. These parameters are the fundamental characteristics of the trapping center, through which the trap can be identified in different devices. In addition, spectroscopic studies can help to understand the mechanisms of charge trapping and to determine the location of the trapping centers in the device.

Extraction of the trap characteristics from the experimental data requires a theoretical understanding of the trapping process. In the presence of an electric field, the characteristics of the capture and emission process change. The apparent activation energy in this case may significantly differ from the zero-field binding energy of the trap. To determine the position of the trap level with respect to the conduction band accurately, the effect of the perturbing electric field must be taken into account.

In this paper we discuss in detail one of the most commonly encountered manifestations of trapping in AlGaN/GaN transistors: gate lag. The focus of the paper is on the physical properties of the traps, the mechanisms of the charge transfer, and the spatial location of the traps in the device. In Section 2, we present a brief overview typical trapping effects observed in AlGaN/GaN transistors and describe the mechanism of gate lag. Section 3 discusses application of the transient current spectroscopy (TCS) as the trap characterization method in AlGaN/GaN HEMTs grown by plasma-assisted molecular beam epitaxy (MBE). In Section 4, we present a detailed experimental investigation of the
charge carrier capture and emission from the traps. The analysis allows one to determine the physical characteristics and the location of the active traps in the device. Finally, we discuss the origin of traps in AlGaN/GaN transistors in Section 5.

2. Trapping effects in GaN transistors

Trapping behavior in GaN-based FETs has been recently reviewed by Binari et al. [24]. The most commonly encountered and usually the most pronounced effect in the AlGaN/GaN HEMTs is gate lag. It reflects recharging of the trapping centers as a result of variation of the gate potential. Charge temporally trapped in the vicinity of the transistor channel can reduce the drain current level by as much as 90% [25]. In general, the trapping centers can be located on the device surface, in the AlGaN barrier, or in the GaN buffer. Because of a strong correlation of the effect with the semiconductor surface treatment, it was concluded that at least some trapping centers are located on the surface [15, 18, 26]. It is believed that the AlGaN surface contains a large density of ionized donor states [22]. The gate lag therefore has been associated with the ionized donor states located on the surface between gate and drain electrodes [26]. The temporal character of charge emission from these traps is typically a stretched exponent with a characteristic time in the range of seconds [26, 27]. Practically no quantitative investigation of these trapping centers exists because of difficulties of the analysis of the stretched exponent dynamics. The presence of the trapped charge on the surface was confirmed by scanning Kelvin probe microscopy [27]. The measurements showed that electrons migrate up to 0.5–1 µm along the surface away from the gate contact.

The surface states, however, are not the only source of the gate lag. The trapping centers in the barrier or in the buffer also affect the density of the 2DEG. The barrier trapping occurs due to charge tunneling from the gate into the semiconductor. The tunneling is assisted by a strong transverse electric field across the gate-to-channel barrier layer. The field also enhances the charge emission from the barrier traps. The field effects, therefore, are particularly important for the barrier traps and they must be taken into account during the characterization. The characteristic times of the field-assisted emission may vary from hundreds of nanosecond to milliseconds. Traps located in the buffer are usually associated with current collapse and drain lag [28]. The transient effects related to these traps however appear in the gate lag measurements. The bulk traps were found to be sensitive to illumination and the information about the energetic location of the trapping levels was obtained from the photoionization spectroscopy [3, 29]. The spectrum revealed two broad absorptions corresponding to the traps in the middle of the GaN bandgap [29–31].

The non-exponential character of the trapped charge emission, observed by many groups, complicates quantitative characterization of the defects. Models of broad spectrum of trapping states have therefore been proposed to explain this behavior. On the other hand, DLTS studies of the defects in GaN and AlGaN Schottky diodes, which are solely sensitive to the trapping centers in the bulk, show distinctive spectral signatures. It implies that the trapping centers are characterized by localized levels within the bandgap. Most commonly observed are the deep levels with activation energies 0.18–0.25, 0.4, and 0.6 eV [18, 21, 32–34]. A few studies showed similar DLTS peaks in the HEMT geometries
confirming the localized nature of the observed traps [15, 16]. Spectral broadening can occur due to large densities of the defect states. The wavefunctions of the trapped electrons in this case overlap causing electron delocalization and formation of a miniband. The non-exponential character can also be caused by a non-uniform electric field distribution in the sample, emission from several different trapping centers, and phonon coupling.

2.1. Mechanisms of gate lag

Gate lag is a delayed response of the drain current with respect to the gate voltage variation. Consider a system of equivalent localized trapping centers in the vicinity of the gate contact, with the ground level of trap within the bandgap. The potential at the gate electrode defines the position of the trap levels with respect to the Fermi level and therefore, its variation causes changes in the occupation factor of the trapping center. An electron can be captured on the trap from the conduction band or from the gate electrode. The occupation factor \( f_T \) is described by the balance of the capture and emission processes

\[
\frac{df_T}{dt} = C_{\text{tun}}(1 - f_T) + c(1 - f_T) \frac{n}{N} - ef_T \left(1 - \frac{n}{N}\right). \tag{1}
\]

Here, the first term represents electron tunneling from the gate into the semiconductor. The other two terms represent electron exchange between the trapping level and the conduction band, where \( c \) and \( e \) are the capture and the emission probabilities, and \( n/N \) is the occupation factor of the conduction band. In the barrier, the occupation factor is very small \((n/N \ll 1)\).

The charge dynamics is derived from Eq. (1). In the equilibrium, the emission and capture processes balance each other resulting in the steady state occupation factor \( f^0_T \). When a negative potential is applied to the gate, the probability for electrons to tunnel through the gate contact barrier increases significantly. The additional flow of electrons results in an increase of capture coefficient \( C_{\text{tun}} \). The occupation factor rises and reaches a new equilibrium state \( f^0_T + f^*_T \). The capture dynamics can be obtained from Eq. (1) by neglecting \( n/N \) terms

\[
f_T(t) = f^0_T + f^*_T(1 - e^{-C_{\text{tun}}t}). \tag{2}
\]

The inverse of the capture coefficient \((C_{\text{tun}})^{-1}\) represents the characteristic time of the capture process.

When the negative gate potential is removed, the filling process is interrupted. The non-equilibrium trapped charge, however, temporally remains localized on the defect level. The system returns to its original state with the emission as the dominant process. The corresponding transient dynamics can be approximated by an exponential function

\[
f_T(t) = f^0_T + f^*_T e^{-et} \tag{3}
\]

with the characteristic time \( e^{-1} \). Eqs. (2) and (3) show that the discrete levels produce exponential results.

The density of the 2DEG is affected by the electric field of the trapped charge. The dynamics of the trapped charge is, therefore, directly reflected in the channel current. It will be shown later, that for the source-drain bias below the knee voltage, a small deviation
of the channel current from the steady state value is directly proportional to the amount of trapped charge. The channel current response to a gate voltage variation exhibits two stages: an instantaneous change of the current to an intermediate level followed by a gradual approach to a new steady state level. The latter corresponds to charge migration in and out of the traps.

3. Transient current spectroscopy of the traps

The emission and capture rates \( e \) and \( C_{\text{tun}} \) can be measured by monitoring the time evolution of the drain current. The emission probability depends on the temperature and the position of the trap level with respect to the bottom of the conduction band. Various forms of transient spectroscopy are based on the measurement of the functional temperature dependence of the emission and capture rates, from which the activation energy of the trap and its capture cross-section can be extracted [35]. Most widely used are the transient current spectroscopy (TCS) and transient gate capacitance spectroscopy (also known as deep level transient spectroscopy (DLTS)). Both methods have distinct advantages and, therefore, complement each other. TCS is sensitive to the trapping effects throughout the device structure. However, it is often difficult to isolate a particular defect in the presence of several different active trapping centers distributed within the device structure. The transient capacitance spectroscopy, on the other hand, addresses only the trapping centers directly under the gate. However, DLTS is less sensitive than TCS because of the small gate capacitance in actual devices.

3.1. Transient current spectroscopy

At room temperature, the emission process from defects in GaN is typically thermally activated. Consider a simplified model of an electron localized on a level inside the bandgap. The electron can acquire sufficient thermal energy to overcome the trap potential barrier and escape from the trap. The thermal emission probability is derived using the principle of the detailed equilibrium 

\[
e(T) = AT^2 \exp \left(-\frac{E_A}{kT} \right)
\]

where \( E_A \) is the activation energy of the trapped charge and \( A \) is a constant. In the simplest case, the activation corresponds to the position of the trap level with respect to the bottom of the conduction band \( E_A = E_T \) for a donor-type trap (or with respect to the top of the valence band for an acceptor-type trap). The constant \( A \) is related to the capture cross-section of the trap \( \sigma \):

\[
\sigma = \sqrt{\frac{2\pi}{3}} \frac{A\pi \hbar^3}{m^*k^2}.
\]

Both characteristics of the trap can be found by fitting the temperature dependence of Eq. (4) to the experimentally measured emission rate. A trapping center can be unambiguously identified by the activation energy and the capture cross-section.
Dynamics of the trapped charge is directly reflected in a deviation of channel current from the steady state level. To establish the relationship between the trapped charge and the channel current, we consider the AlGaN/GaN heterostructure transistor as a parallel plate capacitor with the gate contact as one electrode and the 2DEG as the other. As we mentioned earlier the charge can be trapped either on the surface and/or in the barrier. First, consider the barrier traps. A negative charge placed between the gate and the channel induces a compensating positive charge at the electrodes. The total amount of the induced charge equals that of the trapped charge. The distribution between the electrodes depends on the location of the trapped charge. Assuming that the induced charge in the transistor channel is much smaller than the total channel electron density, it can be approximated by a simple expression \( \Delta q_{2\text{DEG}} = -Q_T \left(1 - \frac{d}{d_{2\text{DEG}}}ight) \), where \( Q_T \) is the trapped charge, \( d \) and \( d_{2\text{DEG}} \) are the barrier thickness and the distance between the trapped charge and the channel. The closer the trapped charge to the channel the stronger effect it has on the channel electron density. If the charge is trapped on the open surface rather than in the barrier, the amount of the induced charge in the channel equals exactly the trapped charge \( \Delta q_{2\text{DEG}} = -Q_T \), since all the trapped charge must be compensated by the channel. In both cases the induced channel charge is proportional to the amount of the trapped charge \( Q_T \). The carrier mobility in the channel remains unchanged for relatively small variation of the carrier density in the channel. The trapped charge, therefore, is directly proportional to the difference between the steady state current and the transient current \( Q_T(t) \propto \Delta I(t) = I_{SS} - I_D(t) \).

3.2. Current transients in GaN/AlGaN/GaN HEMTs

A schematic diagram of the GaN/AlGaN/GaN HEMT used in our study is shown in Fig. 1. The heterostructure is grown by plasma assisted molecular beam epitaxy (MBE) on semi-insulating 6H-SiC. A 60 nm thick AlN nucleation layer is first deposited at a substrate temperature of 830 °C. The nucleation layer is followed by approximately 2 µm of undoped GaN grown at 0.5 µm h⁻¹ with a Ga flux just below the transition to Ga droplet formation. The substrate temperature is 745 °C. The GaN buffer is followed with a
30 nm thick Al$_{0.34}$Ga$_{0.66}$N barrier and the heterostructure is completed with a 5 nm GaN capping layer. The substrate temperature is not changed during the deposition of the barrier structure. The upper 15 nm of the AlGaN barrier and the 5 nm GaN capping layer in some devices are doped with Si at a level of $1 \times 10^{18}$ cm$^{-3}$. Devices with Si doping typically show the best RF power performance and exhibit less pronounced gate lag [36, 37].

The devices are fabricated using optical contact lithography. After dry etch mesa isolation, ohmic contacts are evaporated with drain-source openings of 5 µm. The Ti/Al/Ni/Au ohmic metal stack is alloyed at 780–850 °C in N$_2$ atmosphere to form a good contact to the 2DEG. Lastly, 1 µm long Schottky gates are deposited by e-beam evaporation of Ni (300 Å) followed by Au (3000 Å). The chips are not passivated before measurement. Each HEMT consists of two opposed gate fingers, with total gate periphery ranging from 50 to 200 µm.

The transient current measurements are realized with the device held at a constant source-drain bias in the common-source configuration. A steady state current $I_{DS}^{SS}(V_G)$ is flowing in the channel. To fill the trapping centers with electrons, the gate voltage is switched from the high level $V_{G}^{SS}$ to a lower level $V_{G}^{P}$ for the duration $\tau_p$. The channel current drops in the response to the gate pulse. In the same time, electrons from the gate electrode start tunneling into the semiconductor and filling available trap states. When the gate potential is switched back to the initial level, the channel current recovers only an intermediate level. The charge trapped during the filling pulse partially depletes the channel and limits the current level. The difference between the current level after the filling pulse and the steady state level corresponds to the number of the trapped electrons and the current transient represents the dynamics of charge emission from the traps. The source-drain current transient is measured using a low insertion impedance 100 MHz bandwidth current probe.¹

Fig. 2 shows two limiting cases of the channel current response typically observed in our unpassivated devices with Si doping. In this experiment the transistor is pinched off most of the time ($t < 0$) and all the available trapping centers are filled. At $t = 0$ the gate potential is switched to the on state ($V_G = 0$ V) for 10 µs. One of the devices in Fig. 2 shows an instantaneous current recovery, while the other exhibits obvious gate lag. After the initial current switching to $\sim 85\%$ of the steady state level, the drain current slowly completes the full recovery within 50–100 µs. Typical devices with Si doping exhibit $\sim 90–95\%$ initial recovery. The characteristic times of full recovery are similar for doped and undoped devices [38].

The rate of current recovery increases at elevated temperatures. An illustrative example of the temperature dependence is shown in Fig. 3(a) for an undoped sample. The series of normalized transients were measured at temperatures ranging from 283 to 363 K. Prior to the measurement, the device is held under the source-drain bias in the pinch off state ($V_D = 12$ V; $V_G = −11$ V) for $\sim 10$ ms. During this period, the number of captured

¹ Tektronix A6312. We avoid measuring the current using a load resistor. The transient change of the channel resistance produces variation of the actual source-drain voltage drop. As a result the channel recovery increases; it only slightly affects the temporal dynamics in the case of small transients. However if the emission rate critically depends on the applied field it can result in a faster initial transient, which slowly approaches the actual emission rate.
electrons saturates. As the gate potential switches to $V_G = 0$ V, the captured electrons slowly emit from the traps. The corresponding channel current transients exhibit long exponential tails allowing accurate measurement of the electron emission rate from the traps. The variation of the emission rate with the temperature is consistent with the thermal emission mechanism. Fig. 3(b) shows $eT^{-2}$ plotted against the inverse temperature. The activation energy of the process $E_A$ is found to be $0.22 \pm 0.01$ eV. The capture cross-section is $6.7 \pm 0.7 \times 10^{-19}$ cm$^{-2}$ [38]. The measured activation energy however does not always correspond to the binding energy of the electron on the trap. Later we address the effects of the electric field in the structure, which can significantly change the apparent activation energy.

It is obvious from the shape of the transients that the dynamics of the trapped charge is more complex than a single-exponential decay. In Fig. 3(a), the transient contains two distinctive stages with different characteristic times. Only the latter dynamics follows the exponential decay law. In fact, it is not always feasible to isolate the exponential tail. A typical approach curve exhibits a non-exponential character suggesting that the model presented earlier is oversimplified.

There are few factors that can result in non-exponential character of the transient: (i) Electrons are trapped on several discrete trap levels, in which case the transient is a sum over exponential decays with different rates and amplitudes. (ii) The trapping centers form a continuous distribution of energy levels and the electrons are emitted from all the levels.

---

2 These transients were obtained in devices where the top layers of the structure were not doped with Si. In general, we observed larger amplitudes $\Delta I_D$ compared to the doped devices.
in the trap band. The transient character in this case is rather a stretched exponent. (iii) If the emission process is assisted by an electric field, the non-uniform field distribution in the structure results in variation of the emission rate spatially. The overall apparent emission rate in this case slows down as the electrons first escape from the traps located in the high-field region.

3.3. Selective probing of the trap states

Gate lag is often caused by several different trapping centers. The transient in this case appears as non-exponential and extraction of the emission rate becomes ambiguous. The emission rate for each level can be measured by means of selective probing. In general, the probability of capturing an electron under applied negative gate voltage varies for different traps. By tailoring the depth and width of the gate filling pulse, the single trapping centers therefore can be selectively activated [39].

Fig. 4 shows an example of the drain current transient in an unpassivated GaN/AlGaN/GaN HEMT, where two types of the trapping centers are reflected. As the voltage switches from $V_G^p = -7$ V to the on state $V_G = 0$ V, the current instantaneously rises to $\sim 95\%$ of the steady state level. Then the current level reaches $\sim 99\%$ within a period of a few microseconds. This dynamics corresponds to the charge emission from the fast state. It is followed by a much slower process that continues for hundreds of microseconds. The transient indicates the presence of two traps with significantly different emission rates.
The fast portion of the transient, however, has a non-exponential form and the precise value of the emission rate is difficult to extract. Noting that the amplitude of the fast transient is relatively large, we reduce the depth of the filling pulse. The response of the channel current to the short (500 ns) and shallow ($V_{G}^{P} = -3$ V) filling gate pulse shows that the state with a fast emission rate is still activated, while the transient due to the slow trap is negligible. The inset on the left of Fig. 4 shows the difference current $\Delta I(t)$ normalized to the saturation value $I_{SS}$ for $V_{G}^{P} = -3$ V. The population of the traps decreases exponentially and the characteristic time of $\sim 1$ µs is easily found by fitting $\Delta I(t)$ with an exponential function. As the depth and duration of the filling pulse increases the slow dynamics becomes more pronounced. The right inset in Fig. 4 shows $\Delta I(t)$ for $V_{G}^{P} = -10$ V, which has the exponential character as well. The characteristic time of this process is larger by two orders of magnitude. If the duration of the filling pulse is extended to 0.1–1 ms, the transient amplitude increases and the character becomes non-exponential.

4. Analysis of the trapping processes

4.1. Electron capture by the traps

Selective trap filling by means of control of the width and the depth of the filling gate pulse adds considerable value and flexibility to the spectroscopic measurements. To understand further the capture process we discuss effects of the filling pulse parameters on the recovery transient. The rate of emission from the traps is not affected by the initial occupation factor and, therefore, by the filling pulse parameters. Typically, we observe a small variation ($<20\%$) in the emission rate as the depth of the filling pulse increases from
The amount of the trapped electrons, and therefore the amplitude of the current transient, critically depends on the filling pulse parameters. During the filling pulse, electrons from the gate are migrating through the Schottky gate contact, the barrier height of which is $\sim 1.0–1.6$ eV. For deeper filling pulses, the field across the barrier is stronger and the tunneling probability is larger. Therefore the transient amplitude increases with the filling pulse depth. The amplitude also depends on the duration of the filling pulse as expected from Eq. (2).

Fig. 5 shows the normalized channel current for a series of filling pulse with the pulse duration $\tau_p$ ranging from 20 ns to 100 $\mu$s. The amplitude of the transient $\Delta I_D$, outlined by the dashed line in the plot, increases with the pulse duration until it saturates after $\sim 50$ $\mu$s. The curve reflects the dynamics of the filling process, which is close to the exponential character of Eq. (2). The characteristic time of the process is $\sim 10$ $\mu$s. The amplitude of the transient is displayed in Fig. 6(a) for $T = 300$ and $T = 200$ K. The line shapes practically overlap showing no temperature dependence of the capture process.

The amplitude of the transient $\Delta I(t = 0)$ is shown in Fig. 6(b) as a function of the filling pulse depth. Efficient filling of the trap states starts only for the sufficiently deep gate pulses, when a large electric field substantially tilts the barrier band structure. As the depth increases the number of the trapped electrons rapidly increases first, then it slows down near the pinch off voltage. At this point the channel under the gate becomes depleted and an additional increase in the applied gate voltage results only in minor band tilting. The transient amplitude also increases with the drain voltage for a given pulse depth. It suggests that the capture process is enhanced by the applied field.
As in the case with the duration of the filling pulse, the shape of the transient amplitude in Fig. 6(b) is independent of the temperature. We conclude therefore that the leading mechanism by which the electrons migrate from the gate electrode to the traps is the direct tunneling. The electric field assists the tunneling process and results in the large number of the trapped electrons in the vicinity of the gate. The characteristic time of the process seems to be independent of the applied field.

4.2. Field-assisted emission from the traps

Analysis of transient current spectroscopy requires detailed understanding of the emission process. The activation energy $E_A$ extracted from the temperature dependence of the emission rate is the energy that a localized electron needs to acquire to overcome the barrier of the trap. In general, this energy can be different from the trap level position with respect to the bottom of the conduction band. Traps characterized by a repulsive long range potential are one example. The activation energy in this case is larger in the amount of the repulsive barrier height. Underestimation of the trap level, on the other hand, occurs if the trapping center is subject to an external electric field, which lowers the trap barrier in the direction of the field vector. This case is particularly important for AlGaN/GaN HEMTs, where strong fields exist in the barrier of the structure. Here, we address the effect of the electric field on the emission rate and on the apparent activation energy.

In the presence of the high electric field, the trap potential barrier height can be substantially lowered as shown in Fig. 7, causing an increase of the electron emission probability. This effect, known as the Poole–Frenkel effect, has a distinctive functional
Fig. 7. Schematic diagram of the attractive electrostatic trap potential in the presence of the applied electric field. Arrows represent three possible mechanisms of emission from the trap: thermally activated emission over the lowered barrier due to the Poole–Frenkel effect (PF), phonon-assisted tunneling (PAT), and direct tunneling (DT).

dependence on the field strength. The trap barrier decreases in the amount \( \Delta \phi_{PF} \) proportional to the square root of the electric field \( F \) (for a Coulombic-type trap)

\[
\Delta \phi_{PF} = \left( \frac{q^3}{\pi \varepsilon} \right)^{1/2} F^{1/2},
\]

where \( q \) is a unit of electron charge, and \( \varepsilon \) is the dielectric constant of the material [40]. The corresponding activation energy of the trap becomes field dependent \( E_A(F) = E_A(0) - \sqrt{q^3 F / \pi \varepsilon} \), where \( E_A(0) = E_T \) is the binding energy of the electron on the trap in the zero field. The expression suggests that the activation energy of the traps located in the region of a high electric field (10^6 V cm\(^{-1}\)) can be up to 0.2–0.25 eV smaller than the zero-field binding energy. The emission process from the trap is, therefore, strongly enhanced by the field with the emission rate \( e(F) = e(0) \exp(\Delta \phi_{PF} / kT) \) increasing exponentially with the square root of the field.

An example of the Poole–Frenkel emission from the traps in GaN is shown in Fig. 8(a), where the emission rate is plotted as a function of the potential difference between the gate and the drain terminals [39]. The characteristic emission time rapidly increases from a few milliseconds at low fields (\( V_D = 2.5 \) V) to sub-microsecond at higher fields (\( V_D = 7–8 \) V). To verify the functional dependence, the measured values of the emission rate are fitted with a power law function (\( \ln e = a + bV^p \)). The result of the fitting \( p = 0.53 \) suggests that the emission rate increases exponentially with the square root of the applied field confirming the PF behavior. The solid line in the plot shows a fit to the data \( e = e(0) \exp(\alpha \sqrt{V_D}) \), where the zero-field emission rate \( e(0) = 0.04 \pm 0.03 \) s\(^{-1}\) and the geometrical factor \( \alpha = 6.4 \pm 0.4 \) V\(^{-1/2}\).

The PF effect has a substantial impact on the activation energy of the trap. The apparent activation energy extracted from the thermal dependence of the emission rate at \( V_D = 3 \) V is only 0.11 \pm 0.01 eV (Fig. 8(b)). However the measured value differs substantially from the zero-field activation energy, which can be estimated using the fitting parameters of both
the field and the temperature dependence. Assuming that the pre-exponential factor $A$ is not modified by the applied field, we find $E_A(0) = kT \ln[e(0)/AT^2] = 0.39 \pm 0.03$ eV. This estimate is based on extrapolation of the field dependence to $F = 0$ and it is critically dependent on the accuracy of the constants $A$ and $e(0)$. In the presence of a strong electric field, the electrons can escape from the trap via alternative processes: the direct or the phonon-assisted tunneling into the conduction band [41]. The mechanisms are schematically shown in Fig. 8. If the tunneling probability is comparable with the thermal emission, the extracted activation energy $E_A$ and the constant $A$ appear smaller than the actual characteristics.

To verify the validity of the PF model the temperature dependence of the emission rate must be measured for different bias conditions. According to the PF effect, the activation energy of the emission process decreases with the applied field. Fig. 9 shows the emission rate for another device as a function of the inverse temperature measured at voltages varying from $V_D = 4.25$ to $V_D = 5.75$ V. In the temperature range of 250 to 360 K, the emission rate follows the classical Arrhenius behavior (Eq. (4)) for all bias conditions. The extracted activation energy decreases with the applied field from $0.14 \pm 0.005$ eV at $V_D = 4.25$ V to $0.089 \pm 0.005$ eV at $V_D = 5.75$ V (Fig. 9, inset) corresponding to the PF trap barrier lowering. The pre-exponential factor $A = 7 \pm 1$ s$^{-1}$ K$^{-2}$ remains constant at lower fields and it increases slightly to the level of $10 \pm 2$ s$^{-1}$ K$^{-2}$ at $V_D = 5.75$ V. As the temperature decreases below 200 K the emission rate becomes temperature independent. This behavior can be attributed either to the presence of the competing emission mechanisms or to the device self-heating.

Fig. 8. (a) The emission rate plotted vs. the square root of the drain voltage $V_D$ for three separate devices. The traps are filled using a 350 ns gate pulse $V_G^P = -3$ V, after which the gate is kept at $V_G = 0$ V. (b) Variation of the emission rate with temperature, shown as $eT^{-2}$ vs. $1/T$ for $V_D = 3$ V.
The results in Fig. 9 show that the electron emission from the trap is thermally activated at temperatures above 250 K. The emission rate in this region must be consistently described by the expression:

\[ e(T, F) = AT^2 \exp \left[ -\frac{E_T - \Delta\phi_{PF}(F)}{kT} \right] \]  

The binding energy \( E_T \) can be determined according to the following procedure:

(i) The pre-exponential constant \( A \) and the apparent activation energy \( E_A \) are estimated from the temperature variation of the emission rate at constant bias conditions (Fig. 9) (in the case that the pre-exponential factor \( A \) depends on the field \( F \), Eq. (7) cannot be used for description of the emission process).

(ii) The activation energy is the difference between the binding energy \( E_T \) and the PF barrier lowering \( \Delta\phi_{PF}(F) \). The latter is extracted from the field dependence of the emission rate. Assuming that the emission rate exponentially increases with the square root of the applied field \( e = e(0) \exp(\alpha\sqrt{V_D}) \) we extrapolate \( \Delta\phi_{PF}(F) \) to
Fig. 10. The measured emission rate (symbols) as a function of applied drain voltage measured at different temperatures. The solid lines show the rate calculated using the PF model (Eq. (6)). For $T = 200$ K and $T = 150$ K, the PF emission rate (dashed lines) is too small to describe the experimentally measured values.

$F = 0$ and find the zero-field activation energy $E_T$. The binding energy $E_T$ is a constant, therefore the sum of the apparent activation energy $E_A$ and the PF barrier lowering $\Delta \phi_{PF}(F)$ must be constant for all the bias conditions, or equivalently, the estimated activation energy $E_A(V_D) = E_T - kT\alpha(T)\sqrt{V_D}$ must be consistent with the measured values (inset of Fig. 9) at all temperatures.

The zero-field binding energy for the device shown in Fig. 9 $E_T = 0.54 \pm 0.05$ eV. The emission rate calculated according to Eq. (7) with $A = 7$ s$^{-1}$ K$^{-2}$ and $\alpha = 6.8$ V$^{-1/2}$ is shown as solid lines in Fig. 10 for various temperatures. The result overlaps well with the experimentally measured values shown in symbols. We conclude, therefore, that above room temperature, the emission process is thermally activated. It is assisted by the electric field due to the gate-drain potential difference via Poole–Frenkel potential barrier lowering. Below 200 K, the emission rate remains constant at the level too high to be explained by the thermal ionization, indicating the increasing relative efficiency of the tunneling effects or the device self-heating.
5. Discussion

5.1. Trap potential structure, location, and density

Identification of the emission mechanism allows unambiguous determination of the binding energy of the trap. In addition, we can deduce other important information such as trap location and its nature. We showed that electron emission from the 0.54 eV trap is well described by the PF model, which implies that the trap is described by a long range attractive Coulomb potential [42]. Therefore this trapping center is an ionized donor-like defect. Traps with similar activation energies have been observed in DLTS studies on GaN Schottky diodes [31–34]. The origin of this trap is unknown at this point.

The dependence of the emission rate on the applied field is indicative of the spatial location of the traps. The PF effect implies a direct relationship between the emission rate enhancement and the field acting on the trap. The substantial enhancement shown in Figs. 7 and 10 requires that the strength of the electric field \( F = 1–3 \text{ MV cm}^{-1} \). This estimate is slightly higher than the field expected in the barrier directly under the gate terminal. Such a field can only exist near the drain-side edge of the gate contact, where the field is enhanced by the edge singularity. We note that the estimated value of the field is approaching the breakdown value. However, the extent of the high field region is only a few nanometers, which is not enough for an electron to gain sufficient kinetic energy to cause the impact ionization. The gate edge also has the highest probability for electron tunneling from the gate metal into the semiconductor owing to the field singularity. The observed PF effect, therefore, unambiguously identifies the location of the trapping centers: near the drain-edge of the gate contact.

To estimate the density of the occupied traps after the filling pulse, we need to establish a relationship between the change of the channel current and the amount of the trapped charge. The trapped charge \( Q_T \) is proportional to the change in the 2DEG density \( Q_T = \alpha \Delta n \), where \( \alpha = 1 \) for the surface traps and \( \alpha > 1 \) for the traps located under the gate electrode. The 2DEG density in the steady state is \( n \cong 10^{13} \text{ cm}^{-2} \) (\( V_G = 0 \text{ V} \)). In the linear regime, the relative change of the channel current equals the change of the 2DEG density. Therefore a lower bound for the active trap density can be estimated from the amplitude of the current transient. In our devices we observed trap densities of \( Q_T \geq 10^{12} \text{ cm}^{-2} \).

5.2. Correlation of traps with MBE growth conditions

One difficulty with the analysis of trapping behavior in AlGaN/GaN HEMTs has been the wide variety of phenomena observed by different groups. Timescales from nanoseconds to seconds have been observed in different devices. The vast majority of studies have been performed on samples grown by MOCVD. With this technique it is known that growth conditions can dramatically alter the observed behavior associated with bulk GaN traps [28]. It also appears that device performance depends critically on the treatment of the free surface between the gate and drain. Our studies have focused on material grown by MBE and we now make a few general observations.

In general, while MBE grown material certainly does exhibit gate lag, the magnitude of the effect appears to be smaller than that observed for the MOCVD grown structures.
In particular, the MBE grown device performance appears to depend less sensitively on surface preparation. This observation is substantiated by the fact that reasonable power densities can be achieved in MBE grown devices without the use of surface passivation techniques [43]. The reasons for this difference are not understood at present. One parameter that can dramatically alter the quality of MBE growth is the gallium to nitrogen ratio used in the growth of the GaN buffer region [17, 44]. Growth under nitrogen rich conditions has been associated with rough surface morphologies and increased densities of point defects [17]. The increased rate of formation of point defects may have a serious impact on the observed trapping behavior. Conversely, while growth under Ga rich conditions leads to smooth surface morphologies and higher electron mobilities, any excess Ga accumulated on the surface can alter the electrical nature of threading dislocations, leading to increased reverse-biased gate leakage [45]. In our system, the best films are always grown just below the transition to Ga accumulation on the film surface. This places a very narrow window for optimal growth by MBE. To our knowledge, no systematic study of the influence of Ga surface coverage on gate lag phenomena has been performed. In addition, Si doping of the barrier and capping layers seems to partially mitigate the effect of traps in our devices [38]. While MBE holds promise, at this juncture, it is premature to claim that any specific trapping behaviors are found in material grown by one technique and not the other.

6. Conclusion

Understanding the mechanisms of gate lag is important for the optimization of the performance and reliability in GaN-based devices. We reviewed the phenomenon in AlGaN/GaN HEMTs. The major origin of gate lag in these devices is related to electron trapping by the states located on the semiconductor surface and in the transistor barrier. Under the influence of the electric field, electrons tunnel through the gate contact barrier into the semiconductor. The electrons are captured by the traps in the vicinity of the gate edge, causing a partial depletion of the 2DEG in the transistor channel.

Identification of the traps in AlGaN/GaN HEMTs and their origin is a critical issue. The physical characteristics of the trapping centers as well as their density and location inside the device structure can be deduced using transient current spectroscopy. The technique also allows investigation of the trapping mechanisms. Transient current spectroscopy is particularly valuable because the characterization is performed on actual devices. While the technique has limitations, it provides important information allowing identification of the individual traps, even in the presence of several trapping mechanisms. Substantial help in understanding of the physics of particular traps in GaN can be provided by other characterization techniques.

Significant research effort is currently directed on trap elimination in GaN-based devices. Careful control of the epilayer growth conditions and surface passivation seem to be the most promising solutions for AlGaN/GaN HEMTs. Modification of the transistor structure design may also be beneficial. Investigations of gate lag as well as other trapping effects provide insight into the trap elimination problem. With a better understanding of
the basic material properties and continuing improvement of its quality, we expect that superior characteristics of GaN will be fully realized.

References


