Invited paper

Review of bias-temperature instabilities at the III-N/dielectric interface

C. Ostermaier\textsuperscript{a,}\textsuperscript{*}, P. Lagger\textsuperscript{b}, M. Reiner\textsuperscript{a}, D. Pogany\textsuperscript{b}

\textsuperscript{a} Infineon Technologies Austria AG, Siemensstraße 2, 9500 Villach, Austria
\textsuperscript{b} TU Wien, Institute of Solid State Electronics, Floragasse 4, 1040 Vienna, Austria


corresponding author.
E-mail address: clemens.ostermaier@infineon.com (C. Ostermaier).

1. Introduction

The advantages of III-N electron power devices originate from the improved material parameters. In particular, the larger bandgap of 3.4 eV in GaN increases the critical electric field. Furthermore, the typical lateral heterostructure design of a barrier layer above a channel layer enables a high carrier mobility within the two-dimensional electron gas (2DEG). Such high electron mobility transistors (HEMTs), consisting of e.g. an AlGaN/GaN structure, benefit from the reduced gate-drain capacitance. The formation of the 2DEG is based on the polar nature of III-N compounds [1–3], creating a potential drop over the AlGaN barrier, and the existence of surface states [4–6], providing the necessary carriers to form the channel. Even though the benefits of such a lateral channel are indisputable, it also creates a clear dependency on the charge states of the surface (or interface in the case of surface passivation). Furthermore, gate dielectrics are an effective way to eliminate gate leakage currents in HEMTs [7]. Such metal-insulator-semiconductor (MIS)-HEMTs would be an ideal choice of device with good controllability of the channel. Several concepts have been introduced to achieve the desired normally-off operation with MIS-HEMTs by either partially [8] or fully recessing the barrier [9–11], or introducing negative charges in the barrier [12] or the dielectric layer [13,14]. Although such normally-off devices show some clear advantages over the alternative junction-FET-type p-GaN-gate devices [15,16], they all suffer from a dramatic forward bias drift or positive bias temperature instability (PBTI). This effect is dominantly caused by negative charge trapping at or close to the III-N dielectric interface. Even though this interface may vary from one type of device to another, depending on the barrier (material, recess, surface condition, etc.) as well as the type of dielectric, quite similar results have been shown regarding the nature of the problem. Moreover, the dielectric interface in the gate stack is from a material perspective identical to the interface between III-N surface and the passivation in the access region of the device. The typical bias conditions in this region causes an electric potential similar to negative gate bias at the gate. Therefore, charge trapping similar to PBTI in the access region is not an issue, even though electron trapping in general due to lateral injection has become a well discussed issue with respect to transient drain current degradation e.g. current collapse or transient on-resistance degradation [17,18].

The forward gate bias drift has been investigated for several different dielectric materials over the past decade using Si\textsubscript{3}N\textsubscript{4} deposited by LPCVD [19–22], ALD [23,24], or in-situ grown Si\textsubscript{3}N\textsubscript{4} [25–29], Al\textsubscript{2}O\textsubscript{3} [23,24,30–37], HfO\textsubscript{2} [19,36,38,39], and SiO\textsubscript{2} [40–44]. Furthermore, AlN as an interlayer between the dielectric layer has been investigated [45–47]. In order to achieve normally-off behavior, several concepts of recessed barrier MIS-HEMTs have been shown [10,48] with qualitatively and quantitatively comparable behavior to none recessed AlGaN/GaN HEMTs [23,49–51].

Ibbetson et al. [4] first introduced the generally accepted idea of electrons arising from donor-like surface states at the AlGaN surface [52]. Early on, it was assumed that the donor state density could be a single defect level basically pinning the surface potential. Medjdoub et al. suggested that the donor density should be limited to $2.6 \times 10^{12}$ cm$^{-2}$ by investigating the surface potential on thick lattice-matched InAlN barriers [53]. Similar results were achieved later by Jiao...
et al. for different Indium contents [54]. Studies on AlGaN barriers have shown lower donor state densities and distributed surface densities of about $10^{13}$ cm$^{-2}$ eV$^{-1}$ [55,56]. It was shown that the surface potential increases with increasing Al content. However, due to the lattice mismatch of the AlGaN barrier on the GaN buffer, the maximum thickness of the barrier that can be studied is limited.

Photo-assisted CV measurements on 30 nm thick Si$_3$N$_4$/GaN MIS stacks which are able to ionize deep donor states at the interface as well as defects related to the dielectric have shown a shift of $\approx 3$ V [57]. We can estimate that this value corresponds to an additional interface charge of unionized donor states of about $4 \times 10^{12}$ cm$^{-2}$, which supports further the idea of a limited number of donor states at the interface. The dependence of the surface barrier height on the surface oxidation was systematically studied by Alomari et al. for different oxidation times in InAlN [58]. While initial oxidation did not modify the surface potential, the formation of a 3 nm oxide layer clearly decreased the 2DEG density beyond the thickness reduction of the barrier. For the AlGaN surface it was shown that already annealing associated with oxidation of Al leads to an increase of the surface potential [59]. From a microscopic point of view, Miao et al. have shown that the bare surface should not obtain any surface donor states in the upper part of the band gap excluding clean AlGaN surfaces as the source of electrons for the 2DEG [60]. They have further studied a variety of oxidized surface structures that could lead to donor states in the upper half of the band gap [61]. The comparison to experimental data on oxidized AlGaN surfaces [59] has indicated the possibility of several oxide surface structures [62,63] to fit well to the experimentally observed surface potential. In addition to the role of elemental oxygen in creating surface states [64], also H$_2$O and OH-bonds were suggested to play an important role in creating a surface state shown by electrochemical investigations [65].

Even though those investigations give some general possibilities on the atomicistic view of the surface, the definite origin of the 2DEG and the interaction with defect in the adjacent dielectric are still largely unknown. The density of atoms on the (0001) surface of GaN is $1.13 \times 10^{15}$ cm$^{-2}$ [2]. The ionized interface donor state density as well as the typical interface state density is in the order to $10^{13}$ cm$^{-2}$, being about 1% of the whole surface atoms. However, the resolution of typical surface analysis techniques is limited to about 1% of the investigated near-surface atoms. Therefore, clear evidence to connect a surface mechanism with a surface contamination is hard to establish.

Attempts to clean or modify the surface have been shown by wet chemical cleaning [66–68], thermal treatments [69,70], or exposure to plasma [71–73]. Even though superficial oxygen amounts can be lowered, a complete removal of the surface states has not been shown so far. Besides oxygen, other adsorbates are also suggested to play a role in electrical device behavior [74–77]. The halogens chlorine and fluorine show a great affinity to III-N surfaces [75,78,79]. Chlorine is suggested to saturate and hence passivate the dangling bonds at the AlGaN surface [75,78], by occupying a bond and thus inhibiting the formation of a surface oxide. However, no significant change in device behavior is observed. Chemically, fluorine will behave in a comparable manner at the III-N surface. In contrast to chlorine, we found that a fluorine termination does significantly change the capture and emission behavior of the donor states in AlGaN/GaN MIS-HEMT devices [77,80].

Dielectric surface passivation has been shown by low-temperature atomic layer deposition, plasma-assisted deposition, as well as high temperature low-pressure chemical vapor deposition (LPCVD), close to the temperature of GaN surface decomposition around 700–800 °C. In situ passivation directly after growth allows an even higher process temperature due to the insignificant time between epilayer growth and surface passivation. Crystalline and sharp interfaces have been shown by ALD as well as LPCVD processes using TEM, claiming to be superior compared to amorphous dielectric materials. However, it is still quite challenging to process contamination-free interfaces (e.g. free of any foreign atoms to either material of the interface) due to the limited cleaning possibilities of III-N surface and the typical base pressures of some milli-torr in classical deposition tools. Hence it is a question of interest how one would need to fabricate gate dielectrics on the very inert III-N surfaces in order to achieve interface qualities known from advanced silicon technologies.

Thermal oxidation of GaN would lead to $\beta$-Ga$_2$O$_3$ with a bandgap reported in the range of 4.4–4.8 eV and a conduction band offset of 0.5 eV to GaN [63,81]. This might be not sufficient in order to suppress gate leakage currents. Therefore, thermally grown $\beta$-Ga$_2$O$_3$ is most probably not a proper candidate as dielectric in general. One additional drawback arises from the columnar growth structure in GaN material grown on foreign substrates. The threading dislocation and also the grain boundaries lead to chemically weaker areas of the surface. Hence, those areas lead to preferential decomposition and oxidation, causing 3D like surface structures and surface roughening [70,82].

From a technological point of view, the GaN MIS interface differs quite strongly from the well-known Si/SiO$_2$ interface which is thermally grown and is formed out of the bulk material [83–87]. Even though several studies on GaN MIS stacks have studied the effect of surface cleaning as well as different ways of post deposition treatments, a defect passivation scheme similar to the passivation of $P_x$ centers at the Si/ SiO$_2$ interface with H [88] has not been found until now.

In this paper, we review the phenomena and effects seen in III-N MIS structures from a defect point of view and define a minimum stability requirement from an application perspective. Section 2 covers some basic background information including the fundamental defect model considered throughout the paper, the basics of III-N MIS stacks, and different measurement techniques used to adequately characterize PBTI and the lifetime conditions. In Section 3 we summarize our main findings from different dielectric materials and introduce a pragmatic lifetime criterion for III-N power devices. Section 4 discusses the role of donor states after surface modification by fluorine which triggers some fundamental questions about the role of donor states in GaN devices. Many authors still consider donor charges as fixed charges [89–92]. Even though this view reflects the typical observations, it is still very much simplified, as ionized defect states even with long capture time should not be considered as fixed charges. Those aspects will be mainly discussed in Section 5.

As the detailed microscopic nature of any involved PBTI-related defects is still unclear, they are often categorized as interface states, border traps [93,94], oxide traps etc. which is often simply based on the applied analysis or evaluation method. In our discussions, we try to refrain from any speculation in this regard and simply refer to any modulating charge states at the III-N/dielectric interface as interface states, charges or defects. Those charge states include all states that are able to communicate with free charges within the semiconductor (charge reservoir). Thus, we consider defect states at the interface or within the dielectric. We continue to believe that any further distinction is not helpful for the discussion unless clear evidence of the nature of the defect is provided.

The paper does not include a discussion of technological aspects to improve the interface as the authors do not believe that an objective relevant understanding exists today. That does not mean that certain improvements are not possible, but it is our understanding that none of the presented results are sufficient to fulfill a lifetime criteria like the one introduced in Section 3. All of the results were shown on AlGaN/GaN HEMT structures without recessing the barrier. Nevertheless, we believe that our findings summarized here are general enough and are representative of PBTI effects with and without barrier recess [50]. Therefore, we focus on the larger picture of the physical aspects of the defects. Obviously, the findings and mechanism seen today in III-N PBTI are closely linked to the understanding that evolved on silicon technologies [95].
2. Models and methods

The following part will define the defect types that were considered for further discussion and integrate those defects into the gate stack with a dielectric layer, the interface to the barrier, and the barrier itself. Their impact on the threshold voltage ($V_{Th}$) is discussed. Finally, we discuss the different methods and their limitations when applied for drift and defect characterization.

2.1. Considerations on defect models

Even though there is a lot of detailed literature on defects at semiconductor/dielectric interfaces [96,97] and on the dynamic behavior of charge trapping and de-trapping [95,98] we would like to define here a few general aspects of defects and how they are considered in this review. We believe this information is mainly helpful in avoiding misunderstandings in the following discussion. The word defect in the most general point of view relates to a localized electronic state (a type of bond structure) which can exchange carriers with a carrier reservoir (here 2DEG). In the simplest picture, the defect can capture an electron from the 2DEG with a capture time constant $\tau_{\text{capture}}$.

$$c_a = 1/\tau_{\text{capture}} = n_0 \sigma_a \sigma_{\text{th}},$$

where $\sigma_a$ is the capture cross-section, $n$ the concentration of available carriers for trapping at the interface and $\sigma_{\text{th}}$ the thermal velocity of trapped carriers. For the emission of electron from the defect back to 2DEG we get

$$e_a = 1/\tau_{\text{emission}} = N_c \sigma_a \sigma_{\text{th}} \exp(-\Delta E_{\text{Trap}}/kT),$$

with $N_c$ being the effective density of states in III-N and $\Delta E_{\text{Trap}}$ the defect activation energy, see Fig. 1a.

A simple example of a defect would be an unpassivated surface bond with at least one charge state within the III-N band gap. However, much more complex defect configurations can exist, particularly in amorphous layers. Due to many degrees of freedom in amorphous materials, such defect types have a broad range of energy levels and capture cross-sections as well as spatial positions within the dielectric. However, as we are often not able to distinguish between defects close to the interface and at the interface we should assume that they can be placed at different positions away from the interface.

An often-used empirical model [99,100] to describe this distribution of defects over different energies $E$ and different positions $x$ within the dielectric layer is a Gaussian model written as

$$N_{\text{int}}(E,x) = N_{\text{int,0}} \exp\left(\frac{E - \mu_{\text{int}}(x)}{2\sigma_{\text{int}}}\right).$$

where $N_{\text{int,0}}$ is the total amount of defects, $\sigma_{\text{int}}$ the standard deviation of the distribution over different energies and $\mu_{\text{int}}$ the peak energy value of the distribution. The spatial character of the distribution inside the dielectric material is represented by the $x$-dependency of $\mu_{\text{int}}$, which is an empirical trick to translate the tunneling effect into different energy positions.

We can therefore formulate the threshold voltage drift during the recovery (i.e. emission) as the sum of contributions from all defects with their respective time constants $\tau_{\text{emission}}$:

$$\Delta V_{Th} = \frac{q}{C_{\text{Dielectric}}} \sum_i N(i) \times e^{\tau_{\text{emission}}/\tau(i)},$$

and for stress (i.e. capture) with the time constants $\tau_{\text{capture}}$,

$$\Delta V_{Th} = \frac{q}{C_{\text{Dielectric}}} \sum_i N(i) \times \left[1 - e^{\tau_{\text{capture}}/\tau(i)}\right].$$

where $N(i)$ is the density of the $i$-th defect type, $C_{\text{Dielectric}}$ represents the dielectric capacitance neglecting that the defects may have different distances to the gate electrode and $q$ being the elementary charge. For simplicity, we have considered that the recovery starts from a state with fully occupied defects and the stress from fully empty defects i.e. the same $N(i)$ in Eqs. (4) and (5).

It has been shown that the interaction between the defect and charge reservoir in the case of interface or bulk oxide states is not only dependent on the electronic Coulomb-like potential but also on the local dynamics of atomic vibrations nearby the defect [95,96,101,102]. The change of the charge state of the defect can change the local atomic arrangements (i.e. lattice relaxation) and can also change its vibrational properties. In consequence, the transition energies between defect and conduction (or valence) band have not only an electronic component ($E_{\text{Trap}}$ level with activation energy $\Delta E_{\text{Trap}}$ Fig. 1a) but also a vibronic one (localized phonon mode represented by parabolic potential of harmonic oscillator). This can be pictured in Fig. 1b by a configuration coordinate diagram [102] where e.g. the total energy of an empty defect state with a free electron at the minimum of the conduction band $E_{c}$ is represent by the parabola "empty" and the energy of the occupied defect state with a parabola "occupied". The equilibrium position of the defect in either state is at the minimum of the parabola. In order to capture an electron to the empty defect, the defect has to acquire the energy $\Delta E_{\text{p}}$ from the lattice. The capture cross-section will be thus temperature dependent according to $\alpha_a = \sigma_{\text{p}} \exp(-\Delta E_{\text{p}}/kT)$, with $\sigma_{\text{p}}$ being a temperature independent prefactor. Consequently the apparent measured thermal activation energy for emission will be (see Fig. 1b), $\Delta E_{\text{Trap,apparent}} = \Delta E_{\text{Trap}} + \Delta E_{\text{p}}$ and not just $\Delta E_{\text{Trap}}$ as considered by many authors. The relaxation to the ground state will occur by emission of phonons. That is why we speak on non-radiative multi-phonon (NMP) relaxation model [102] both for capture and emission processes. Both emission and capture processes could be written by a general equation:

$$\tau(i) = \tau_{0}(i) \times e^{\frac{E_{a}(i)}{kT}},$$

with $E_{a}(i)$ being an activation energy for an $i$-th defect type and $\tau_{0}(i)$ the corresponding prefactor. Thus, for the capture process it holds $E_{a} = \Delta E_{\text{p}}$ and for the emission process $E_{a} = \Delta E_{\text{Trap,apparent}}$. Another consequence of the lattice relaxation is that the optically and thermally induced transitions will have different energies (Frank-Condon shift) [102].

Furthermore, it has been shown that defects at the Si/SiO$_2$ interface and oxide defects can have several different stable configurations for either charge state, where NMP capture/emission processes can occur at transitions between either state [95,101]. Let us consider the generalized situation of Fig. 1b where, for simplicity, the defect in the occupied state can have two stable configurations 1 and 2 represented by corresponding parabolas in the configuration coordinate diagram in Fig. 1c. If the electron is in state 1, it can be emitted to the 2DEG with an emission rate $e_1$ or it can be transferred to state 2 with the transition rate $e_{1\rightarrow 2}$ (energy barrier $\Delta E_{1\rightarrow 2}$). The back transformation from state 2 to 1 is governed by the transition rate $e_{2\rightarrow 1}$ and energy barrier...
Fig. 2 shows the typical lateral structure of a III-N HEMT device. A barrier on top of a GaN channel forms the 2DEG below the interface which is connected by metallic source and drain contacts. The whole surface is passivated providing a defined interface condition. The gate dielectric is typically a separate layer deposited just under the gate, but can in principle be also the same as the passivation layer. It has been argued that the opening of the passivation for the purpose of depositing the gate dielectric can have a deteriorating effect on the interface. However, wet etching or remote-plasma etching have shown very low-damage results due to the inertness of the (0001) III-N surface [103–106]. The stack shows that the III-N/dielectric interface is not just of high relevance under the gate, but also in the whole access region. However, forward bias is only applied at the gate under typical device operations. The access region towards the drain is typically being depleted during off-state or back-biasing conditions.

The band structure through the gate stack is schematically shown in Fig. 3a. Neglecting the influence of the gate metal workfunction (as in the access region) and the buffer charges, the interface charge $Q_{\text{interface}}$, the total amount of polarization charges in barrier/channel part and the 2DEG density $n_s$ must fulfill the charge neutrality. As the barrier does not contribute to any net polarization charges, the total polarization charge equals the one of GaN $Q_{\text{Pol, GaN}}$. Thus, we can write

$$Q_{\text{interface}} = Q_{s, GaN} - q \times n_s.$$

where both terms $Q_{s, GaN}$ and $q \times n_s$ are negative resulting in a positive $Q_{\text{interface}}$. Therefore, we can follow that the interface (as well as the free III-N surface) has to contain dominantly donor-like defect states, which can contribute their electron to the 2DEG.

Hence, in thermal equilibrium conditions (Fig. 3a) the charge neutrality level (CNL) [107], at which the interface/surface is neutrally charged, is above the Fermi level. These charges could in principle exist at the interface and within the passivation or gate dielectric. Some have argued that those defects might even be placed on top of the passivation, assuming that all defect states at the interface have been passivated [108]. We try to answer this question with our results from chapter 4. Besides the dominant charges at/near the interface, charge states can also exist in the barrier and the channel region, in particular closer to the buffer. As long as we do not consider gate bias conditions below $V_{\text{th}}$, the 2DEG will not be depleted, with no trapping effect in the buffer. Therefore, we can neglect any traps below the 2DEG. The barrier/GaN interface is epitaxially grown and by nature does not contain significant defects other than pre-existing threading dislocations from the buffer [109]. Barrier defects have been reported to play a role in the current conduction through the barrier in Schottky contacts [110]. As the barrier is in series to the charge trapping interface, it certainly must play a role in the process. Whether this role is dominant over charge trapping or not can strongly depend on the technology of the barrier as well as the dielectric.

One general aspect of the HEMT structure is the remote interface separated from the 2DEG as shown in Fig. 3a. Quite contrary to typical
MOS structures the effect of interface charges is therefore purely electrostatic unless for very thin barrier layers, where the carrier scattering at the interface charge can reduce the carrier mobility in the channel [111,112]. Hence, any transfer characteristics with drifted threshold voltage exhibit a perfectly parallel shift without distortion. Thus, measuring the transient drain current and translating it into the shift of the threshold voltage ($V_{Th}$) does not require any additional assumptions.

The threshold voltage for the whole stack can be expressed as

$$V_{Th} = -\frac{1}{q} \left[ \Delta E_{C(i-th-j-th)} + \Delta E_{C(BARRIER-CHANNEL)} + \Delta E_{C(BARRIER-DIELECTRIC)} \right]$$

$$- q \Phi_{GATE-DIELECTRIC} + (\varepsilon_{CHANNEL} + \varepsilon_B\sigma_B) \varepsilon_D + (\varepsilon_{CHANNEL} + \varepsilon_D\sigma_B) \varepsilon_B$$

where $\varepsilon$ are the polarization charges [3] at the interfaces of the corresponding layers, $\Delta E_{C(i-th-j-th)}$ the conduction band discontinuity between the $i$-th and $j$-th layers, $\Phi_{GATE-DIELECTRIC}$ the potential difference between the gate metal and the conduction band minimum of the dielectric, and $\varepsilon_B$ the dielectric material, respectively. Alternatively, to the interface charge density $Q_{INTERFACE}$, we can also introduce the interface potential $\Phi_{INTERFACE}$ as the potential of the conduction band of the barrier at the interface above the Fermi level (Fig. 3b). Apparently $\Phi_{INTERFACE}$ is not an independent parameter and cannot be directly measured as it depends on the amount of charges trapped at the interface. However, we can assess $\Phi_{INTERFACE}$ directly from the 2DEG concentration using

$$n_s = \frac{1}{q} \left[ \Delta E_{C(i-th-j-th)} + \Delta E_{C(BARRIER-CHANNEL)} \right]$$

where $n_s$ is known from the threshold voltage $V_{Th}$ and the total gate capacitor $C_{Gate}$

$$n_s = \frac{1}{q} C_{Gate} \times V_{Th} = \frac{1}{q} \varepsilon_{DIELECTRIC} \times V_{Th}$$

In any passivated region e.g. the access region, $n_s$ can also be extracted by Hall measurements allowing the direct use of Eq. (10) to calculate the interface potential.

Under forward gate bias (Fig. 3b) all defect states within the marked area in the dielectric, at the interface, as well as in the barrier, can change their charge state depending on their defect ionization energies, their spatial position, as well as the temperature, and the stress time. This energy range covering all defects that can potentially contribute to interface charges is typically referred to as the active energy region [95]. Any change in the concentration of interface states $\Delta N_{Int}$ will thus cause the corresponding change in threshold voltage

$$\Delta V_{Th} = \Delta N_{Int} \times q \times \varepsilon_{DIELECTRIC}$$

Furthermore, we notice that the trapping of electrons around the interface region during stress causes an increase of the interface potential, thus acting against the gate bias stress. Hence with progressing time the active area is becoming smaller as shown in Fig. 3c. This is accompanied by the re-emission of the already captured electrons, i.e. the initially occupied states lying below Fermi-level start to appear above $E_F$ with a tendency to lose electrons [42]. Therefore, we have to be very careful when we investigate the amount of defects captured at different stress times and biases as the effective stress seen by the interface will depend on the amount of pre-existing trapped charges. This effect was previously described as Coulomb effect [42,113,114] and certainly plays a more pronounced role in GaN compared to other technologies due to the use of a dual-layer gate e.g. barrier and dielectric, and the rather large amount of interface states. The Coulomb effect was also shown to lead to 2nd-order dynamical effects [42].

Fig. 3c and Eq. (11b) also indicate that the amount of threshold voltage drift caused by an additional electron trapped at the interface is strongly depending on the dielectric thickness and the dielectric constant. For very large gate capacitors the threshold voltage drift could be strongly reduced for the same number of trapped defects, which suggests that the permittivity is an interesting design parameter for GaN MIS-HEMT devices.

### 2.3. Measurement techniques

We would like to point out that there are two main objectives under which interface measurement techniques and defect characterization methods can be contemplated. From a pragmatic device performance point of view a measurement technique is required to give an absolute or at least relative measure of the amount of unwanted device variation e.g. threshold voltage shift $\Delta V_{Th}$. Furthermore, in order to investigate the nature of the defects underlying the drift we would like to learn particular characteristics of a defect or a group of defects, like time constants of the capture and emissions process, capture cross-sections, energy barriers and their variation. While the first objective is needed to evaluate different technologies with respect to applications and lifetime criteria, the second one is needed to acquire deep physical understanding in order to prepare new solutions and ideas to avoid or minimize the amount of the respective defects. The first objective can be achieved by electrical techniques [42,95] while the second one requires physical methods to study the individual defects as done by TEM, STM [62], ESR [115], EDMR [116], TDDS [117,118].

The main focus in this review is on the first objective, as there is very little understanding of the microscopic nature of the III-N surface, the formation of defect states of deposited dielectrics on the III-N surfacet, as well as the donor states of the III-N surface itself. Therefore, we will not look into the detailed physical defect model, and rather use a pragmatic defect model. Therein every neutral defect that is in the active area (during stress) will be charged after $\tau_{capture}$ and every charged defect that is outside of the active area (during recovery) will emit its electron after $\tau_{emission}$. This defect behavior can most easily be represented in a $\tau$-space also known as capture-emission-time (CET) map [119,120] shown in Fig. 4. The y-axis represents the capture time, the x-axis the emission time and the z-axis or the color-coding.

Fig. 4. Schematic $\tau$-space (CET map) of a defect with broad distributed capture and emission times. The white line represents the slice that can be captured during C-V measurements and the gray box in the bottom right corner indicates the fraction of the $\tau$-space that is represented in a hysteresis measurement (neglecting the different voltages during the bias sweep).
represents the density of defects that are within a range of capture and emission time (usually one decade). The details of CET map construction from the stress-recovery data are given in [119,120]. As such the \( \tau \)-space only represents the defect density of a certain stress bias. All defects can be described by their density \( N_{\text{int}} \) within this 3-dimensional space of \( \tau_{\text{capture}}, \tau_{\text{emission}} \), and \( V_{\text{Stress}} \) at a fixed temperature. This description becomes insufficient if an already captured electron moves from one point in the map to another due to 2nd-order effects [42]. Typically, one should consider that the range of interest of the \( \tau \)-space should range from the shortest timings in applications e.g. switching times to the total lifetime of a product.

2.3.1. Hysteresis measurements

The time-dependent shift in threshold voltage \( \Delta V_{\text{Th}} \) can be measured from the shift of I-V or C-V characteristics. As shown in Eqs. (4), (5), and (11b), the \( V_{\text{Th}} \) shift is directly related to total concentration \( \Delta N_{\text{int}} \) of negatively charged defects.

Even though hysteresis measurements have often been used to characterize the interface, typical sweep times of several seconds limit the amount of information that can be extracted from such a simple measurement. Results represent therefore only a small part of the \( \tau \)-space (see the narrow rectangle in Fig. 4). Furthermore the sweeping of the gate bias from 0 V to \( V_{\text{Gate,max}} \) and back to 0 V, completely mixes up the effect of bias-acceleration and stress time at every bias condition. Therefore, such a test by I-V or C-V serves only as an initial disaster check but does not reveal any insight into the real dynamics of the threshold voltage drift. A very effective method to separate all those individual stress parameters is the extended measurement-stress-measurement (eMSM) technique [112].

2.3.2. Transient measurement techniques

The principle of MSM techniques is depicted in Figs. 5 and 6 [42,121,122]. During the stress period a positive bias is applied to the gate causing charging of the interface and oxide (border) states. To monitor the degree of \( \Delta V_{\text{Th}} \) during recovery, the drain current \( I_{\text{Drain}} \) or the drain bias \( V_{\text{Drain}} \) of the transistor (in series to a load resistor, Fig. 5) at \( V_{\text{Gate}} > V_{\text{Th}} \), is measured for short moments, which is considered not to influence the trap occupancy [121]. Otherwise, zero bias is applied on the gate during recovery. The MSM technique works particularly well for normally-on devices which allow stressing the device at \( V_{\text{Gate}} > 0 \) V and monitoring the \( \Delta V_{\text{Th}} \) at \( V_{\text{Th}} < V_{\text{Gate}} < 0 \) V. Using a pulsed setup on the gate and drain allows characterization of 100 ns stress pulses on wafer level with a few microseconds of reaction time. Therefore, this method allows scanning a wide range of the CET map.

Some fundamental limitations of the measurement-stress-measurement technique are shown in the Figs. 7–8. First, we assume that the initial measurement to characterize the device, e.g. some transfer characteristics, does not alter the device. In the case of a normally-off MIS-HEMT this is typically given for the gate bias range from \( V_{\text{Th}} \) to 0 V. For gate biases above 0 V, charge trapping at the interface occurs. For \( V_{\text{Gate}} < V_{\text{Th}} \), the potential drop within the buffer layers can lead to additional trapping or detrapping which has to be avoided. For the
actual stress-recovery characterization the device is subjected to stress for a defined time as depicted in Fig. 7a. During the applied stress, charge trapping can occur which increases the threshold voltage. This, however, remains undetected as we cannot monitor it during stress. After \( t = t_{\text{stress}} \) the stress stops and the device starts to recover. Some of the trapped charges may already be emitted prior the first moment at which we can detect the threshold voltage at \( t = t_{\text{response}} \). After repeating the same stress sequence on a new device or on the same device with sufficient recovery, the stress is reapplied for \( t = t_{\text{stress}} \). In case \( t_{\text{stress}} \gg t_{\text{stress}} \), the error of remeasuring the same device several times should be relatively small, even if we assume 2\(^{-\text{nd}}\)-order defect kinetics to be involved (see Section 3.1 and Ref. [42]). From the second recovery trace (see the curve for \( t = t_{\text{stress}} \) in Fig. 7b) we are also only able to recognize the part of the recovery after \( t \geq t_{\text{response}} \). In order to graphically compare the actual drift of the device to the measured one, two points from the recovery traces at \( t = t_{\text{extract}} \) where transferred back to the stress times 1 and 2. The error in this example was chosen to be around 40% from the actual drift behavior. Unfortunately, this error cannot be estimated in real situations, leaving us with an unknown number of hidden defects outside our measurement window.

Those fundamental limitations have further implications when considering the effect of temperature during the measurement, which is important for using temperature-accelerated tests for lifetime estimations [123,124]. In addition to the stress-recovery tests shown in Figs. 7b, 8 includes two more stress curves at elevated temperature (red curves). We assume that the capture of electrons is thermally activated and therefore the slope of the threshold voltage drift is higher compared to the reference temperature. After \( t_{\text{stress}} \), defects start to emit electrons into the conduction band. As the temperature is still elevated, also this process is activated, causing the emission of electrons to start earlier in time and resulting in a steeper slope during recovery. In the given example, we chose the \( V_{\text{th}} \) drift at \( t_{\text{extract}} \) to be identical to the one at reference temperature in order to receive identical stress curves. Even though this example is fictitious, it still describes a realistic scenario that must be considered. Depending on the microscopic behavior of the defects, it might even happen that the temperature-acceleration of recovery is faster than during stress. The resulting apparent \( E_{\text{A}} \) would lead to negative values [99,125]. Fig. 9 shows an example of \( V_{\text{th}} \) drift over temperature ranging from \(-20\) to \(220^\circ\)C with relatively little effect of the temperature. The apparent activation energy calculated from the recovery data after 100 \( \mu\)s of recovery and 10 \( s \) of stress at \( 1-3 V \) were in the range of 0.08–0.13 eV. This is much lower than our reported data using separate temperature control for stress and recovery with values up to 1 eV [124]. For silicon devices, even high values up to 1.5 eV have been reported [14,126]. It should be noted that the extraction of \( E_{\text{A}} \) is usually done for exponential dynamics and not trivial for such recovery data with log-t dynamics i.e. the result of multi-exponential dynamics. As shown by Eq. (6), we consider the \( \Delta V_{\text{th}} \) behavior over time as a multi-exponential trend that cannot be deconvoluted into the individual \( N(i) \).

Even though hysteresis measurements have been extensively used in the past to characterize the interface, it is evident that the at least 2-dimensional \( t \)-space cannot be captured well with a single oscillation frequency that is required to match the capture and emission time (Fig. 4). Therefore, typical CV-based methods are only able to capture a slice of the CET diagram (see line in Fig. 4) which may in certain conditions be still a representative picture. However, these conditions cannot be proven by the CV method itself showing the helpless case of investigations using only such methods. A detailed discussion of this problem can be found in [121]. The determination of \( \Delta N_{\text{int}} \) from CV curve hysteresis is problematic as only some part of the defects react,
which is determined by the probing frequency and the up/down voltage sweeping rate [129].

We would also like to mention that interface states in MISHEMTs were investigated also with photo-capacitance techniques where electrons and holes were generated by above band gap light [27,131], so both types of carriers could be trapped in interface states. In such experiments, the trap occupancy depends on a non-equilibrium electron and hole concentration at the interface as well as the parameters of the defect states (Fig. 1). Due to this undefined distribution of occupied traps and non-existing holes in real MISHEMT operation we do not discuss this technique in more detail. Furthermore, due to the Frank-Condon shift between the optical and thermal energies, interpretation of the extracted density of interface states obtained by techniques based on capacitance measurements under below-band gap photon excitation should be considered with care [130].

Caution has to be also taken when selecting a suitable device size for CV analysis. Since devices with large channel length can exhibit large series resistance effects in the measured capacitance, short channel length devices are preferred for analysis [132]. Similar considerations on sample geometry and substrate resistivity should be taken into account when analyzing vertical MIS capacitor structures.

The conductance-frequency (G-ω) technique is a traditional technique to evaluate the interface defect density and their energy position in the Si/SiO2 system [98], where an expression for weak inversion is assumed. Some researchers have used the same kind of analysis for MISHEMTs. However, there are two striking differences between Si MOSFETs and GaN MISHEMTs: 1) the defect equivalent capacitance CDefect related to Si/SiO2 interface states ($C_{\text{Defect}} = \Delta N_{\text{int}} \times C_p$, where $C_p = \frac{q^2}{kT} = 2 \times 10^{-18}$ F), is much lower than the dielectric capacitance $C_{\text{Dielectric}}$, which is not the case in GaN MISHEMTs where the opposite case is mostly valid. 2) GaN MISHEMTs often include a III-V barrier within the gate capacitance with voltage dependent conductivity. The improper use of the G-ω technique for MISFET analysis in the region of threshold voltage was pinpointed in [129,132]: the conductance peak in this region can be also related to the channel conductivity increase with applied positive bias, and not to interface states as they cannot respond in this regime.

Some authors have used G-ω techniques for $N_{\text{int}}$ evaluation in the spill-over regime where carriers accumulate at the III-N barrier/dielectric interface. The authors of [33] correctly pointed out that G-ω technique cannot be used for defect evaluation in this regime, and they have proposed capacitance-frequency (C-ω) dispersion analysis in spill-over regime as a suitable alternative. In principle, this cannot be true because real and imaginary components of admittance are coupled. Consequently Capriotti et al. have shown that both G-ω and C-ω frequency dispersion techniques are not suitable for defect analysis in MISHEMTs (at least not straightforwardly) [129]. Since the barrier conductivity increases with applied positive bias, this produces a similar frequency dispersion in both G-ω and C-ω curves as the virtual defect response, even if there is no interface state. On the other hand, a large density of interface states (i.e. $C_{\text{Defect}} \gg C_{\text{Dielectric}}$ and $C_{\text{Barrier}}$) will not lead to an increase of the G-ω response but rather cause a saturation. This is because $C_{\text{Defect}}$ is in series with $C_{\text{Barrier}}$ and $C_{\text{Dielectric}}$. Thus, the observed temperature variations of the G-ω curves [33] can be just due to temperature variation of the barrier conductivity, not due to defect response. In general it is difficult to distinguish between the defect and the intrinsic barrier (i.e. without interface defects) response [129].

To be more specific we give an example: Typical values of interface state density obtained in literature by different authors using the G-ω or related techniques are always in the range of 0.5–2 × 10^{11} cm^{-2} [33,39]. The theoretical value of maximum conductance to angular frequency $G/\omega_{\text{max}}$ from [129] is calculated as

$$ (G/\omega)_{\text{max}} = \frac{1}{2} \frac{C_p}{(C_{\text{int}} + C_p)} $$

Taking typical values for $C_{\text{Dielectric}} = 1–6 \times 10^{-7}$ F/cm² and $C_{\text{Barrier}} = 3–6 \times 10^{-7}$ F/cm², one gets a first order estimation for the equivalent density of defects to be $[G/\omega]_{\text{max}} / C_p = 0.7–20 \times 10^{-8} / 2 \times 10^{-18} = 0.4–10 \times 10^{10}$ cm^{-2}, which is of the same order. Although this equivalent trap density could lead to typical values, in reality, this can be only a spurious response of the intrinsic barrier and the CDielectric. Even though the frequency dispersion techniques are not suitable for trap evaluation in MIS-HEMTs they can be used for trap evaluation in true Si MOSFET-like structures where the III-N interface is in direct contact to electron channel [129,133].

From the above, it is evident that the main technique to investigate defects at the interface should include large ranges of stress and recovery times, which makes the classical MSM or OTF the most adequate techniques.

3. PBTI in GaN

This chapter summarizes all relevant effects that have been found in PBTI studies with link to the physical models of Section 2. Furthermore, a comparison of different dielectric materials as well as lifetime considerations for switching applications is given.

Some of those fundamental studies were performed on an AlGaN/GaN MIS-HEMT with deposited TEOS (Tetraethyl orthosilicate); others used SiN as gate dielectric. While some effects seem stronger for one dielectric than for the other we do not claim here that it is related to the dielectric alone. We assume that the total stack of AlGaN barrier and dielectric is responsible for the overall behavior. As some of those studies have been performed on wafers from different types of epitaxial layers and independent runs, we do not engage in a comparative study of those different test devices. Our focus instead is to present the different aspects of electrical behaviors that have been found in all dielectrics and are common to them.

3.1. Interpretation of drift effects

This part discusses the electrical results and understanding of typical MSM data. Fig. 11 shows a typical recovery measurement at a fixed bias and different stress times revealing the typical broad distribution of recovery times. The data clearly indicates that we are not able to speculate about the total amount of defects that might have recovered at much faster times than the reaction time of the measurement setup of 3 μs. Similarly, the maximum recording time (few kiloseconds) is not representative for a permanent drift. Naturally we assume that all trapped electrons would emit at much longer recovery times. One interesting fact from this data can be derived from the slope of the threshold voltage drift per decade of recovery time (Fig. 12). We see that for stress times longer than 100 ms the defect density $N_{\text{int}}$ decreases. This indicates that certain captured electrons transition to a

Fig. 11. Typical recovery behavior of a SiO2 MIS-HEMT at 4 V gate bias and for a series of stress times [42].
different defect state with longer emission times. Such model has been introduced earlier in the concept of multistate defects [95] and also discussed in terms of 2nd-order defect kinetics which causes negative entries in the CET map [42]. The change from concave to rather convex shaped recovery trends can further be imagined by electrons moving into an energetically more stable state. This state could be either lower in energy or deeper into the dielectric (see e.g. state 2 in Fig. 1c). During the recovery, this can apparently increase the emission time since in order for the defect to be able to emit the electron it has to be first reconfigured back to the original state (see state 1 in Fig. 1c).

Another manifestation of this multi-state defect behavior can be seen in repetitive stress experiments [114]. Fig. 13 shows two recovery traces after an equal amount of stress time. One curve was recorded after a continuous stress time of 100 μs. The other one was recorded after 100 times 1 μs stress pulses interrupted by intermediate 100 ns recovery phases. One would expect the continuous stress to at least lead to the same amount of ionized defects or more: the result initially surprises by demonstrating the opposite. However, considering that our measurement window is limited and does not include the whole dynamics of the involved charge states, we believe that the recovery trace after continuous stress of 100 μs was initially equal or above the repetitive 100 × 1 μs stress result. We further assume that traps filled during continuous stress show a shorter emission time on average. After repetitive stress, the average emission time of the ionized defects increases. This hypothesis is supported by a simple consideration of the dynamics of repetitive stress. We assume that any capture of electrons leads to trapping into defect states with a broad range of emission time constants, as suggested by Fig. 11. Considering now that during the short intermediate recovery only the fast emitting defects can be discharged, the average emission time constant of all the involved defects should increase for each re-capturing stress cycle. This effect is initially independent of 2nd-order defect kinetics, which allows moving their electron from one state to another defect state without emission into the charge reservoir. Furthermore, we should consider that in the multistate defect model the transformation barriers between different defect states depend on the applied bias and equilibrium position of each defect state in space. The response of the same defect can be different during the stress and recovery periods for different bias conditions [95]. Besides the action of multi-state defects, the Coulomb effect can lead to similar 2nd-order kinetics effects [42,114]. In order to discriminate between the two effects, careful comparison with simulation would be required.

In Fig. 14 we can see recovery curves for two different stress times and a series of different gate biases. Fig. 15 shows the extracted threshold voltage drift over stress time for different stress biases after fixing the recovery time. It shows that the number of negatively charged defects increases with voltage as well as with stress time. Even though Fig. 15 suggests an almost perfect linear relationship for the defect state density with respect to stress time and stress bias, we should note again that the recovery of those defects strongly differs, as discussed earlier. Two points can be followed from the behavior of the stress conditions: With increasing gate bias the active energy region (Fig. 3b, c) increases linearly causing more defects to be accessible to electron trapping [95]. Second, the gate bias reduces the barrier potential for electrons to be captured [42,134], therefore decreasing the capture time.

The influence of different voltages and related memory effects is further shown in Fig. 16 by gate step-stress. The device has been initially stressed at 3 V (curve with green triangles) and then continued to be stressed at 4 V causing a proportional amount of drift (red
diamonds). Directly after, the device was stressed at 3 V again as indicated by the arrow in Fig. 16 (green circles). The device initially recovered due to emission of electrons from defects that have been within the 4 V-related active energy region, but were outside the 3 V-related region. In addition, the Coulomb effect certainly reduces the effective active energy region under all bias conditions. However, it is very unlikely that it acts as the main mechanism for electron emission at the 2nd 3 V stress sequence. That is because the electron emission starts right after the gate bias step from 4 to 3 V and $\Delta V_{Th}$, proportional to the interface potential, remains constant at that moment. The fact that the threshold voltage increases further for $t > 10^2$ s after a short decrease at $t < 10^2$ s shows that even more defects with longer capture time exist in the 3 V-related active energy region. We assume that the intermediate stress at 4 V has accelerated electron capture into some defects by lowering the respective ionization barrier.

3.2. Barrier-related drift effects

So far, we considered that all the dynamic effects of the threshold voltage during stress and recovery experiments can be attributed to the III-N/dielectric interface. Even though we did not lay out any proof for this assumption, it seems intuitively right given the nature of the dynamics. Bulk traps typically have a less complex dynamic pattern without such a broad distribution of capture and emission times [135]. However, we can recognize one effect seen in some devices that seems to rather relate to the transport of carriers from the 2DEG to the interface than the defect dynamics at the interface itself [134]. Fig. 17 depicts the stress data from two samples with silicon oxide and silicon nitride, respectively.

Considering the linearity of the stress data for different gate biases (Fig. 15), we would assume parallel shifted lines in Fig. 17a as indicated by the solid lines. However, we can recognize that for 1 and 2 V the early onset of the drift rather lacks behind in threshold voltage drift [134]. This effect is seen much stronger in the device shown in Fig. 17b using a different sample with Si$_3$N$_4$ as dielectric. The drift clearly starts at different stress times depending on the gate bias. Investigating the temperature dependency of this on-set of the drift revealed an activation energy which linearly decreases with the gate bias proportional to the potential decrease at the interface (Fig. 18a). The only discontinuity of this linearity was found for gate biases below 1 V. We speculate that this effect could be explained by a field-enhanced process e.g. defect-assisted tunneling transport allowing an enhanced conductivity through the edge of the barrier near the interface (Fig. 18b). A similar mechanism has been suggested for the leakage current model in Schottky gate HEMTs [110,136]. The tunneling mechanism, which appears at high electric fields (here appearing at low biases) apparently decreases the thermal activation energy. The activation energy of the electron transport of 0.65 eV at gate bias of 0 V is comparable to the interface barrier potential of 0.75 eV, extracted from
2DEG density and the material parameters using Eq. (10). This suggests a predominant transport via the AlGaN conduction band in GaN MIS-HEMTs.

We can therefore conclude that in general the capture process is a serial process with an effective capture time constant \( \tau_{\text{capture, effective}} \), being a sum of the capture time constant of the actual defect \( \tau_{\text{capture, defect}} \) and the rate-limiting time constant related to the electron transport process through the AlGaN barrier \( \tau_{\text{Barrier}} \):

\[
\tau_{\text{capture, effective}} = \tau_{\text{capture, defect}} + \tau_{\text{Barrier}}.
\]

Such a barrier time constant is essentially inversely proportional to the leakage current over the barrier \( \Phi_{\text{Interface}} \) (see Figs. 3b and 18b): \( \tau_{\text{Barrier}} \approx \frac{1}{\Phi_{\text{Interface}}} \). We would like to note that even tough in this particular case the transport over the barrier is dominantly conduction band related, transport over dislocation bands can be considered as well [136]. As the conductive property of the III-N barrier can strongly depend on growth and processing conditions, the relevance of \( \tau_{\text{Barrier}} \) on the overall stress-recovery response can thus be technology dependent.

### 3.3. Comparison of different dielectric materials

Up to this point we have tried to discuss the general behavior seen in representative III-N/dielectric stacks. The following part is aiming to compare different dielectric materials in order to show the impact of the dielectric constant on the double dielectric layer stack [19]. Looking at Eq. (8) we would assume that high k dielectrics provide a benefit to a MIS stack by increasing the dielectric capacitance, therefore bringing the interface charges “electrostatically closer” to the gate, and thereby reducing their impact on the channel. E.g. in the case of an infinitely high dielectric constant the interface charge would have no impact on the 2DEG. Fig. 19 shows the threshold voltage drift for different gate biases, for a fixed stress and recovery time [19]. Comparing these data with the stress-time dependent data from Fig. 15, it should be noted that those graphs only represent partial information of the whole CET-map. They describe therefore not the full nature of the defects similar to hysteresis measurement (Fig. 4). Therefore, such a comparison might lack in objectivity compared to the detailed investigation shown in Section 3.1. It should be noted that those stacks include different dielectric materials as well as AlGaN barrier thicknesses. The onset at which the gate bias causes a significant potential drop to start initiating the trapping at the interface may depend on the overall structure. However, in this set of data (Fig. 19) we cannot deconvolute the effect of the electrostatics of the stack and the effect of the capture and emission time distribution of the individual defects. AlO\(_3\) as an example shows almost no drift up to 3.5 V of gate bias, which could in part be explained by defects that almost fully recover for the respective stress times and gate bias up to 3 V. This hypothesis is supported by similar results on fully recessed devices (i.e. without barrier between channel and AlO\(_3\) as gate dielectric) [99]. Therefore, a comparison between different dielectrics by such a method is subjectively based on the chosen stress and testing conditions. In order to provide a comparison, one would need to compare the entire behavior of threshold voltage drift against a common lifetime criterion (as will be discussed in Section 3.4).

In order to remove some parts of this misleading behavior we calculate the interface state density which is proportional to the threshold voltage in comparison to the displacement charge \( Q_{\text{Displacement}} \) defined as

\[
Q_{\text{Displacement}} = C_{\text{Dielectric}} \times V_{\text{Gate}}.
\]

Fig. 20 shows all the dielectric stacks after the normalization, revealing a convergence against the theoretical upper limit. Silicon oxide, having the lowest dielectric constant, has the lowest amount of interface charges, while HSiO\(_4\) with the highest dielectric constant has the highest amount of interface charges \( > 10^{14} \text{cm}^{-2} \) in our comparison. Apparently, this trend is limited by the Coulomb effect as the interface charges cannot exceed the displacement charges. As \( Q_{\text{Displacement}} \) is limited by \( V_{\text{Gate}} \) (Eq. (14)) our electrical detection of interface charges is limited by this maximum as well.

Furthermore, the converging data points in Fig. 20 show an offset to the theoretical upper limit. However, we have to be careful arguing that this offset contains any actual improvements beyond the physical limit of \( Q_{\text{Interface}} \leq Q_{\text{Displacement}} \) due to the intrinsic measurement errors: Only accessible defects within the active energy range are being charged that can fulfill \( \tau_{\text{capture}} < \tau_{\text{stress}} \) and as well as \( \tau_{\text{emission}} < \tau_{\text{recovery}} \). An apparent relative improvement seen between one dielectric and another could therefore simply be the results of different defect dynamics rather than an effective reduction of defect states. It is completely possible that the actual density of interface states is much higher (e.g. \( 10^{14} \text{cm}^{-2} \)) than the density we can measure electrically. The only effect caused by different technological approaches as surface preparation, cleaning, dielectric material, deposition process, post-deposition treatments, etc. might be the modification of the defect parameters (Fig. 1c).

### 3.4. Lifetime requirements

In comparison to low-voltage logic devices, \( V_{\text{Th}} \) shifts in power devices are often a bit less critical due to the increased gate bias margin during operation. From a very pragmatic point of view the gate dielectric needs to fulfill at least the following two common criteria for normally-off operation:

First, a sufficiently low threshold voltage drift has to be assured in order to obtain

\[
\Delta V_{\text{Th,max}} = V_{\text{BOL}}(V_{\text{Gate,use}}) < V_{\text{Gate,use}} - V_{\text{Th},0} - \Delta V_{\text{Th,max}}(V_{\text{DS(on),max}} < V_{\text{on},\text{min}}),
\]

Fig. 19. Threshold voltage drift versus gate bias stress for different dielectric stacks. The AlGaN barrier thickness is 20 nm except for the SiO\(_2\) wafer, which has a 30 nm thick AlGaN barrier [19].

Fig. 20. Normalized representation of Fig. 19 showing the interface charges as a function of the displacement charges. The continuous line represents the maximum possible amount of interface charges to be ionized during the equivalent stress [19].
where $\Delta V_{Th,max,EOL}$ is the maximum $\Delta V_{Th}$, at end-of-life (EOL) caused by the maximum steady-state gate bias in use conditions $V_{Gate,use}$ in order to assure that the maximum specified on-resistance $R_{DSon,max}$ [122] and the minimum specified source-drain current $I_{DSon,min}$ can be achieved at a certain minimum overdrive voltage $V_{OV,min}$. $V_{Th}$ is the initial threshold voltage under equilibrium conditions. Obviously, this is the absolute worst-case requirement neglecting any $V_{Th}$ range defined by the application. Second, the stability of the gate dielectric at $V_{Gate,max}$ must be assured till end-of-life. Even though both criteria refer to steady-state conditions, overshoot spikes during switching have to be considered in addition.

Analyzing a whole switching cycle we know that during on-state stress $\Delta V_{Th}$ increases reaching its maximum at the end of the switching cycle. Therefore, the gate source capacitance $C_{GS}$ is also being slowly charged, thereby reducing the chances of spurious turn-on behavior during turn-off by increasing the ratio of $C_{GS}/C_{GD}$. During off-state stress the device can partially recover from forward gate bias stress, slightly reducing $\Delta V_{Th}$ and therefore also $Q_{GSS}$ during re-turn on.

In order to evaluate $\Delta V_{Th,max,EOL}$ correctly, one should model the most critical AC stress pattern (minimum $t_{recovery}$ and maximum $t_{stress}$) over the expected lifetime. $\Delta V_{Th,max,EOL}$ is then equal to the threshold voltage shift at the end of one additional stress pulse. In principle this needs to be evaluated without any recovery delay. In order to avoid this one can assess the $R_{DSon}$ and $I_{DSon}$ during application tests of a device with EOL gate bias stress similar to OTE measurements.

From those considerations we conclude that the main criteria to be fulfilled is that after the maximum threshold voltage shift at EOL, we require still a sufficient amount of charges in the channel under the gate at $V_{Gate,use}$ to maintain a reasonable $R_{DSon}$ and $I_{DSon}$. We can assume that the minimum channel charge under the gate $n_{ch, min}$ should be close to the 2DEG density $n_{2DEG}$:

$$R_{DSon,max} \rightarrow I_{DSon,min} \Rightarrow n_{ch, min} \approx n_{2DEG}.$$  

The required overdrive voltage is therefore noted as

$$V_{OV,min} = q \cdot n_{2DEG} \cdot C_{Gate}. \quad (17)$$

Inserting the term $V_{OV,min}$ into Eq. (15) and substituting $V_{Gate, use}$ by the dielectric thickness and the maximum critical field, allows us to note down the worst-case criteria for the maximum threshold voltage shift over lifetime as

$$\Delta V_{Th,max} = E_{Dielectric,critical} \times t_{Dielectric} - V_{Th,0} - q \cdot n_{ch, min} \cdot C_{Gate}.$$  

$E_{Dielectric, critical}$ is the typical maximum electric field for lifetime conditions. We can further simplify this equation to derive an upper limit for $\Delta V_{Th, max}$ by setting $V_{Th,0} = 0 V$ (optimum case) and $C_{Gate} = C_{Dielectric}$ (maximum of $C_{Gate}$ as $C_{Gate} \leq C_{Dielectric}$), as it would be the case for fully recessed MIS transistors without a barrier under the gate electrode. Finally, we can derive the upper limit for the maximum interface state density $\Delta N_{Int,max} @ EOL$ based only on the parameters of the dielectric and the channel density, which is a design parameter of the device:

$$\Delta N_{Int,max}@EOL \leq \frac{1}{q} E_{Dielectric, critical} \times t_{Dielectric} - n_{ch, min}.$$  

Table 1 contains the maximum interface state density at end-of-life conditions for several typical dielectrics including the channel density $n_{ch, min}$ (i.e. only the result of the left term on the right side of Eq. (19)). $\Delta N_{Int,max} @ EOL$ is therefore in the range of $4-8 \times 10^{12} \text{ cm}^{-2}$ considering typical channel densities of about $0.5-1 \times 10^{13} \text{ cm}^{-2}$. Even though this satisfies the criteria from Eq. (15), it should be considered that this substantial amount of interface charges also significantly increases the electric field in the gate dielectric. Therefore, the stability of the gate dielectric (criteria 2) needs to be carefully reviewed in conjunction with the internal amount of trapped charges.

We would like to note that this criterion is an absolute worst-case criteria and their might be good reasons to limit $\Delta V_{Th, max}$ to smaller values based on the specific application requirements e.g. compatibility with drivers, etc.

After defining a lifetime criterion, the following part tries to answer the question how much PBTI evaluation under DC stress can show relevant behavior for the more typical operation conditions of AC stress. In contrast to previous results discussed in Section 3.1, we will focus here on 1st-order kinetics effects using devices with less dominant 2nd-order behavior. Further details on the topic can be found in Ref. [137–139].

Typical stress and recovery data for the device that has been used in this analysis are plotted in Fig. 21 for a representative gate bias of 4 V and increasing stress times. The device exhibits a fast recovery in the microsecond regime reaching full recovery for stress times up to 1 s for the defined measurement window. The $\Delta V_{Th}$ starts to increase strongly for stress times > 1 s as shown in the stress-plot for several extraction times (Fig. 21a). In the case of our representative device a typically broad distribution of defects is found with an increase towards higher capture and emission times in the millisecond regime (Fig. 22).

Fig. 23 compares DC-gate bias stress for 100 s with AC-gate bias stress for 100 × 1 s and different intermediate recovery times $t_{recovery}$. It can be recognized that the recovery trace after DC stress shows the highest $\Delta V_{Th}$ and merges with the recovery traces after repetitive stress. The difference to Fig. 13 is explained by 2nd-order kinetics effects (see also Fig. 1c). This behavior can be systematically explained using the concept of the CET map [119,120] mentioned in Section 2.3, Fig. 24 shows schematic CET maps for two different stress and recovery scenarios. The green area represents the defects that are occupied. It can be noted that the measurement limits only the maximum capture time and the minimum emission time. The minimum captures time as well as the maximum recovery time is basically dependent on the energy distribution of defects included in the dielectric. Extrapolating the stress and recovery data of Fig. 21 suggests that we should consider minimum capture times in the range of nano-seconds and maximum emission time well above the typical lifetime for the given stress-recovery analysis.

For a single DC stress pulse (Fig. 24a–c) all defects in the CET map with a capture time up to the maximum stress time will be filled (Fig. 24b, CET map after the stress). During the recovery, those defects with $t_{emission} < t_{recovery}$ will be emitted (Fig. 24c, CET map after recovery). In the case of a repetitive stress (Fig. 24d–f) the initial CET map filling after the first pulse resembles the one of the DC-stress except that only fewer defects with capture times up to the initial stress pulse are filled. After every additional stress pulse more and more defects with longer capture time will be included. At the same time, the overall recovery time for already captured electrons increases. Therefore, the occupied area of the CET map decreases. This behavior can be shown by the examples of the CET map directly after the last stress pulse, where $t_{AC} \times N_{Pulses} = t_{DC}$ (Fig. 24e), and after one additional recovery cycle (Fig. 24f). Thus the final CET map occupancy after one additional recovery cycle is quite similar in the two cases. This difference has to be seen in relative proportion to the overall size of the occupied $\tau$-space and the defect density per area of the CET map. Both depend on the

<table>
<thead>
<tr>
<th>Dielectric</th>
<th>$\varepsilon_{Dielectric}$</th>
<th>$E_{OL,critical}$ (MV/cm)</th>
<th>$\Delta N_{Int,max}@EOL$ and $n_{ch}$ (cm$^{-2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>3.9 $\times$ $e_0$</td>
<td>6</td>
<td>$1.3 \times 10^{12}$</td>
</tr>
<tr>
<td>Si$_3$N$_4$</td>
<td>7 $\times$ $e_0$</td>
<td>3.5</td>
<td>$1.3 \times 10^{12}$</td>
</tr>
<tr>
<td>ONO</td>
<td>4.4 $\times$ $e_0$</td>
<td>6</td>
<td>$1.4 \times 10^{12}$</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>8 $\times$ $e_0$</td>
<td>3.2</td>
<td>$1.4 \times 10^{13}$</td>
</tr>
</tbody>
</table>
material and need to be carefully evaluated for each technology.

However, if we compare the CET map occupancies before the final recovery (Fig. 24b and c), we can see that a significantly larger portion is covered after DC stress compared to AC stress. This difference is responsible for the initial increase of threshold voltage shift after DC stress (Fig. 23). The validity of the CET map model can be further highlighted by simply calculating the drift for the measurement conditions from Fig. 23 by using the CET map data (Fig. 22), similarly as in Ref. [119]. The data shown in Fig. 25 are derived using the scheme discussed earlier, such defect kinetics refer to trapped electrons moving from one occupied state into another occupied state. In terms of the CET map this equals moving from one cell, representing a certain combination of capture/emission time constants, to another cell. Hence, the 1st-order kinetics approach discussed in Figs. 23 and 25 is potentially underestimating the drift by simply calculating the drift for the measurement conditions from Fig. 23 by using the CET map data (Fig. 22), similarly as in Ref. [119]. The data shown in Fig. 25 are derived using the scheme depicted in Fig. 24, integrating all specific defect densities for all $\tau_{\text{capture}}$ and $\tau_{\text{emission}}$ that remain trapped at any given test conditions. The continuous stress curve for 100 s stress is shown as reference. The minor inconsistencies seen by the slight mismatch of the calculated drift data compared to the measured one are caused by 2nd-order kinetics effects which could not be eliminated by the calculations.

Even though the difference shown by the final CET maps of AC and DC stress (Fig. 24c and d) is rather small, it certainly becomes more significant for microseconds stress pulses as well as for lifetime considerations. Fig. 26 shows a theoretical calculation for longer AC stress (10 kHz, 50% duty cycle) and DC stress times above $10^4$ s for a constant defect density CET map i.e. a CET map with a constant defect density per decade of capture and emission time. As mentioned before two additional input parameters are required to define the boundaries of the occupied states in a CET map: the minimum capture time and the maximum emission time of defects which are still having an impact on the drift behavior. Since the minimum stress time impacts the DC and the AC drift by the same offset it can simply be omitted from the calculation (see the arbitrary offset in the y-axis of Fig. 26). The maximum emission time was set to $M \times \tau_{\text{capture}}$ for $M = 6, 8, \text{and 10}$, which roughly fits the recovery data (Fig. 21). The calculation result considering only 1st order defects shows a saturation of the drift at $\tau_{\text{Stress}} = M \times \tau_{\text{Pulse}}$, where $\tau_{\text{Pulse}}$ is the stress duration of a single stress pulse (50 μs in our example). However, an important assumption in this concept of the CET map is that no multi-state defects ($2^{\text{nd}}$ or higher order defect kinetics) are involved in the trapping process [42]. As discussed earlier, such defect kinetics refer to trapped electrons moving from one occupied state into another occupied state. In terms of the CET map this equals moving from one cell, representing a certain $\tau_{\text{capture}}/\tau_{\text{emission}}$ into another cell. Hence, the 1st-order kinetics approach discussed in Figs. 23 and 25 is potentially underestimating the drift if the recovery starts prior to the monitoring scheme. This can be recognized also from our AC-stress data (Fig. 23), where the AC drift merges with the DC drift results even before the respective $\tau_{\text{Recovery}}$. Considering the CET map model (Fig. 24), the AC drift should be even slightly lower at $\tau_{\text{Recovery}}$. Even more extreme results have been discussed in Fig. 13, where the AC stress caused a higher drift at long recovery times compared to the DC stress. Hence, the applicability of the CET map model for lifetime assessment excluding 2nd-order effects needs to be separately evaluated.

Even though we used relatively short stress times in order to demonstrate the 1st-order kinetics of the defect behavior, it should be clear that the concept can be further extended to longer stress times closer to lifetime conditions. Therefore single stress (DC-stress) conditions can be used to provide a lower estimate of the lifetime. The applied bias for this stress should be set to the highest applicable gate bias according to the maximum application conditions as the number of accessible defects is proportional to the active energy region (Figs. 3 and 15). This apparently hinders also the use of voltage-accelerated tests. Unfortunately, also temperature acceleration cannot be applied due to the parallel acceleration of capture and emission processes as discussed in Section 2.3.2. In addition, one has to take care that the reaction time of the setup should be at least close to the shortest off-state time considered in switching applications to further avoid underestimation of the drift.

An additional method for lifetime extrapolation using the MSM technique is the application of temperature acceleration only to the
defect capture processes (i.e. during stress) [123]. This can be achieved by applying gate stress while increasing the temperature to certain elevated values above the reference temperature for a certain time. Before the stress is turned-off and the device is being monitored during recovery, the temperature has to be brought back to the reference value. We have shown the implementation of this measurement scheme for GaN by on-wafer heaters beside test transistors applying around 20 W of heat for local temperature increase of about 200 K [124].

Fig. 27. T-accelerated stress behavior at 3 V gate bias and different stress times by in-situ heater [124].

were extracted after 10 ms of recovery at 25 °C independent of the temperature during stress. Under the assumption that the temperature-independent pre-exponential factor $\tau_0$ (Eq. (6)) is same for all the defect states independent of their energetic and spatial position, we can translate the defect behavior at a certain time $\tau_1$ and at a temperature $T_1$ into a different $\tau_2$ at a temperature $T_2$ using [123]

$$\tau_2 = \tau_0 \times \left(\frac{T_1}{T_0}\right)^\frac{T_1}{T_2}. \quad (20)$$

This is particularly useful when using measurements at elevated temperature $T_{\text{meas}}$ in order to accelerate the drift behavior. Using Eq. (20), the drift behavior at time domain of $\tau_1$ can be extrapolated to time domain of $\tau_2$.

Fig. 24. Schematic of transient pulse of $1 \times 100 \text{s}$ (a) and $100 \times 1 \text{s}$ (b) including the schematic CET map right after the end of the (last) stress pulse (b, e) and after 1 s of additional recovery (c, f). It can be recognized that the area of the CET map is quite similar in both cases explaining the drift result of Fig. 23.

Fig. 25. Calculated drift recovery using the CET map data in Fig. 22 and applying the CET map scheme of Fig. 24. The AC stress sequence follows the identical one used in Fig. 23. The black data points represent the measurement points after DC stress as reference.

Fig. 26. Comparison between AC and DC stress for a constant defect density CET map following the scheme of Fig. 24. The calculation is based on a 10 kHz AC stress with 50% duty cycle.
4. Role of native interface donors

An apparently interesting question is, whether there is a fundamental link between PBTI-related\(^1\) interface defects and the native donor states known to exist due to the formation of the 2DEG. If such link exists, it would directly explain why we would not be able to create stable devices, as donor states have to be known to exist in any GaN HEMT device. However, most authors, including us, do not believe in such a link even though a proof or reasoning is usually not stated.

In this part, we will discuss the threshold voltage drift in devices with a modified interface such that the dominant trapping is caused by the interface donor states responsible for the 2DEG. It is obviously not easy to find a process that would allow such a surface modification. Several authors have shown the possibility of modifying the interface potential by cleaning, plasma etching, thermal treatments or any other surface process [66-73]. We have tried several surface treatments by applying thermal or plasma energy onto the surface using a number of available gases and precursors. Eventually we found a remote plasma fluorination process using NF\(_3\), CF\(_4\) or SF\(_6\) that would lead to a very different device result related to modified surface donors. The plasma was produced by an inductively coupled (ICP) generator on a Matson Aspen II at 900 W and 13.56 MHz. AES and XPS measurement show a surface coverage of up to 16% of F, where the F is in part substituting the nitrogen atom thus increasing the surface electronegativity [140].

Further details on the process and its physical analysis can be found here [80].

Devices were prepared in such a way that the fluorination was applied right before the dielectric LPCVD-Si\(_3\)N\(_4\) passivation. From experiments with different dielectrics e.g. SiO\(_2\) deposited in a PECVD, we could verify that the typical behavior seen after fluorination was not limited to the Si\(_3\)N\(_4\) layer. In contrast to previous reports on surface fluorination [141] our treatment did not involve any impinging ions or ion penetration into the AlGaN layer. Even though SIMS measurements did not show any evidence of additional fluorine being left at interface after surface passivation [80], the electrical behavior clearly indicated a difference in the defect behavior. We therefore concluded that the surface modification induced by fluorine radicals is thermally very stable and does not disappear during the subsequent dielectric deposition.

Standard device characterization by IV measurements (Fig. 29) revealed a more negative device threshold voltage \(V_{th}\) with a peak at about 2 V, which indicated the reduced saturation current \(I_{\text{Drain, sat}}\) for the fluorinated device. Gate capacitance and transconductance did not change. Hall measurements on both devices proved a slightly reduced \(n_\text{s}\) of \(5.5 \times 10^{12} \text{ cm}^{-2}\) after fluorination compared to \(7.3 \times 10^{12} \text{ cm}^{-2}\) in standard devices, explaining the difference in \(I_{\text{Drain, sat}}\).

Fig. 30 shows the typical recovery data of the standard device (a) and the stress and recovery data of the fluorinated-interface device (b and c). Both devices were stressed at different biases from about \(-15\) to \(+12\) V for \(100\) s without any prior stress. While the standard device exhibits the typical PBTI-related drift characteristic for stress (not shown) and recovery, the fluorinated device seems to have already achieved a stable condition (quasi-equilibrium) prior to the stress time of the measurement setup with almost no further increase of \(V_{th}\) during stress (Fig. 30b). The emission time constants seen from the recovery in Fig. 30c seem relatively narrow distributed. Due to the large range of \(V_{th}\) shift it was not sufficient to monitor the devices at a single gate bias. Instead a whole range of gate biases was used in repetitive measurements shown in Fig. 31, allowing the calculation of \(V_{th}\) at each moment during recovery [77]. The comparison between the two devices (Fig. 32) clearly shows that the standard device does not exhibit any measurable drift during negative gate bias and starts drifting only for positive biases. In clear contrast, the threshold voltage of the fluorinated device follows the stress bias in positive and negative direction, suggesting fast emission as well as the capture of electrons at the interface.

The NBTI behavior seen in Fig. 32 suggests a new behavior of the fluorinated device that does not exist in the reference MIS device, in which only trapping under forward gate bias stress is recognized. The defect dynamics observed for negative gate bias in the fluorinated device seems to be extending to forward gate biases suggesting a new type of defect with a common behavior in both bias regimes. From the tightly distributed nature of the defect dynamics with almost exponential relaxation behavior, we assume that the atomic nature of the defects is rather well oriented and the defects should be rather located at the interface than in an amorphous (oxide) structure. Considering the charge neutrality of the gate stack (Eq. (7)) and the quasi-equilibrium 2DEG density (Eq. (11a)) for different gate biases before the onset of the device drift, we can derive the density of states (DOS) of the interface donors (Fig. 33) [77]. The peak of the DOS is further into the bandgap (see the line with symbols in Fig. 33) compared to the equilibrium interface potential of the device without fluorination. The latter one is calculated from the 2DEG density and schematically shown in Fig. 33 by an arbitrary peak shape (see the gray solid line). The increased interface potential for the fluorinated device is in line with the reduced saturation current as well as with the reduced carrier density from Hall measurements. We can also note that the donor states induced by fluorine do not pin the interface potential (strongly), as seen by minor changes in the interface potential depending on the amount of trapped electrons. Furthermore, we recognize that for gate voltages below \(-15\) V all defect states seem to be ionized. Therefore, we can conclude that the fluorination acted in one of two possible ways (Fig. 34): Either

---

\(^1\) Even though the term PBTI does not define any particular type of defect, we are using PBTI-related defects here to refer to the kind of defect behavior described in Section 3.
the native donors have been modified into the apparent donor defects after fluorination, or the native donor states have been passivated or shifted even deeper into the bandgap and new donor states have been created. No additional defect behavior appeared under our testing conditions. Therefore, we would exclude further defects deeper in the bandgap within the detectability of our measurement window (> 1000 s). In comparison to the native donor defects, we also assume that the defects caused by fluorination have a larger defect capture cross-section.

The investigation of the microscopic nature of these defect states after fluorination has led to following hypothesis [80]: The most intuitive is that the native states must have been modified either directly or indirectly by fluorination. A direct modification is explained by a substitution of the near-surface ad-atoms by fluorine (be it fluoride, fluorine radicals or similar species at this point). Even though no additional fluorine has been found at the fluorinated interface to the dielectric in comparison to a non-fluorinated interface by SIMS measurements, this finding does not allow any conclusion about its chemical potential and bonding state. The fluorine plasma species might lead to energetically and atomistic different structures and states when binding to the surface, even if the final fluorine concentration after dielectric deposition remains unchanged. An indirect modification is explained by the catalytic activation or modification of the surface. The superficial fluorine, bound to the metals atom at the surface region, might cause a reconstruction or catalyzes another near surface reaction when leaving the surface. Either way, this newly found bonding structure is required to be stable under further processing of the dielectric layers [80].

The investigation of the microscopic nature of these defect states after fluorination has led to following hypothesis [80]: The most intuitive is that the native states must have been modified either directly or indirectly by fluorination. A direct modification is explained by a substitution of the near-surface ad-atoms by fluorine (be it fluoride, fluorine radicals or similar species at this point). Even though no additional fluorine has been found at the fluorinated interface to the dielectric in comparison to a non-fluorinated interface by SIMS measurements, this finding does not allow any conclusion about its chemical potential and bonding state. The fluorine plasma species might lead to energetically and atomistic different structures and states when binding to the surface, even if the final fluorine concentration after dielectric deposition remains unchanged. An indirect modification is explained by the catalytic activation or modification of the surface. The superficial fluorine, bound to the metals atom at the surface region, might cause a reconstruction or catalyzes another near surface reaction when leaving the surface. Either way, this newly found bonding structure is required to be stable under further processing of the dielectric layers [80].

Even though the fluorination process provides a well-defined drift behavior, it would not be accurate to claim that the originally seen PBTI related defects would not exist anymore in such devices. The reason being that the capture process after fluorination is extremely fast and occurs prior to the response time (≈ 3 μs) of our setup (Fig. 30a).
Therefore, the pragmatic view on the drift results with respect to a lifetime criterion seems currently more useful. The main criterion seems to be the limit of increase of the $R_{\text{DSon}}$ due to large $V_{\text{Th}}$ shifts (Eq. (15)) limiting the overdrive voltage. The derived figure of merit for different dielectric materials (Eq. (19)) shows that the maximum defect density should be $< 10^{15} \text{cm}^{-2}$ depending on the 2DEG density. A higher defect density at end-of-life testing conditions would cause an increase in on-resistance or a decrease in saturation drain current. As there are very few results today that can be used to roughly estimate the lifetime conditions for a dielectric interface, we assume that there are currently no known solutions to achieve a sufficiently low defect density. Our own results summarized in Fig. 20 suggest that the only limitation for the number of electrons trapped at the interface (for sufficiently high bias, long stress time, and short recovery time) is the electrical detection limit given by the gate bias. In other words, we assume that the typical drift behavior of $\Delta V_{\text{Th}}$ in GaN MOS devices is simple converging at the level of the gate bias as shown in Fig. 28 for an extrapolation beyond the lifetime of 10 years. Several authors have claimed to have shown improved dielectric gate stacks on fully, partially, or none re- cessed barrier stacks with almost no hysteresis. Yonehara et al. pre- sented recently a recessed SiO$_2$ GaN device with no recognizable drift after 1000 s forward gate bias stress at 150 °C. Hopefully a more de- tailed PBTI study will be shown soon to clarify whether this results is an absolute improvement against the lifetime requirements. Unfortunately many reports have shown insufficient evidence for an absolute im- provement of the dielectric layer or the interface. Insufficient mea- surements are typically characterized by too short stress times (seconds rather than hours), long recovery times (seconds rather than $\mu$s or ms), and limited gate bias ($V_{\text{Gate,max}} < V_{\text{Gate,use}}$). Therefore, a transition to test conditions closer to lifetime conditions would be highly appre- ciated to bring the discussion of GaN interfaces to a new level.

As a result it is still not clear if the state-of-the-art achievements regarding the applied technological methods to improve the interface are sufficient. One reason is certainly the inertness of III-N materials in combination with the limited thermal budget that can be applied to GaN in comparison to SiC. Early decomposition of the GaN surface at relatively low temperatures $> 700$ °C is one of the challenges in pro- ducing native oxide on GaN. Deposited dielectrics have to overcome the challenge of producing a low defect interface on top of a relatively inert surface. We speculate that the relatively large drift results seen nowa- days independent of the dielectric could be a hint that the dominant defect type is situated in a native III-N surface layer. We do not ne- cessarily consider this layer to be a complete monolayer. A fraction of a monolayer e.g. 1% would in fact be sufficient to completely dominate the electric results beyond the defined lifetime criteria. In this case the differences in transient drift behavior seen for different dielectrics can originate from differences in the band alignment and variations of the interface potential. Several reports have argued to have achieved a crystalline interface shown by TEM using various methods of deposition [22]. However, we claim that the accuracy of the TEM is not sufficient to proof a 1% interface defect, as even the best instruments require at least a few tens of atoms per atomic column to represent a single atom. Therefore, the required crystallinity needed from an electrical point of view would require a far higher resolution.

Even though there is currently no proof for the hypothesis that a certain (acceptor-like) defect state exists at the GaN surface, which would dominate the whole PBTI-related drift effect independent of the deposited dielectric, we also do not see it currently as falsified either by any of the available results. As mentioned earlier, it is still not clearly identified whether the PBTI-related defects seen in MIS devices are linked to the native donor states. Our argument for the separation of the detailed parameter of such interface defects are single defect characteristics like in random-telegraph-noise (RTN) or time- dependent defect spectroscopy (TDDS) [95,117]. A first result on GaN MIS devices has proven the feasibility of applying this method on nm- sized devices [118].

Hence, any positive gate bias increase is immediately compensated by charge trapping at the fluorine-induced defect. In other words, the F- related donor states react much faster than the PBTI-related defects, masking thus the possible PBTI response. We do not believe that fluorine-induced defect states are interacting with PBTI-related defects, due to the absence of $2^{nd}$-order trapping effects. Therefore, we conclude that the F-related donor states are likely to be spatially separated from the PBTI-related defects.

5. Synthesis

The results from Section 3 suggest several physical mechanisms to be involved in our drift dynamics. Fig. 35 summarizes those physical aspects that were proven to exist in typical devices for different dielectric layers and interfaces. Unfortunately, the series of drift data for different gate biases, stress times, recovery times, and even temperatures (separated for stress and recovery process [124]) creates still an ambiguous set of data that does not allow us to clearly extract in- dividual defect parameters. As an example, in order to model the dif- ferent energy levels of the trap states correctly (Fig. 1b, c), we would need to know their exact position in the dielectric layer in order to take into account the probability of an electron transition between the conduction band and the defect. Therefore, we want to clarify that we can only argue about the existence of those effects but we are not able to specify any detailed defect parameters of our involved defect states. We have to consider that all the assumptions that we need to make in order to extract a more detailed picture, have a defining characteristic on the outcome of our model. Even Eq. (3) as shown in Fig. 35 already carries more implicit assumptions than we are able to correctly verify. Based on the learning from the Si/SiO$_2$ interface, the only method to
those two types of defects is based on the different dynamic behavior. Let us consider an interface (Fig. 3a) reduced to a single (or distributed) donor-like defect type as depicted in Fig. 3b, where the same defect is responsible for the formation of the 2DEG as well as for trapping during forward gate bias conditions. In this assumption however, all donor defects ND would also contribute to negative VTh drift, which contradicts all our results in Section 3. The F-modified donor state presented in Section 4, however, falls exactly into this behavior shown in Fig. 36. Therefore, we believe in the coexistence of donor states defining the 2DEG density and additional interface defects causing electron trapping during forward gate bias stress (i.e. "PBTI" defects). Hence the nature of those PBTI defects would be acceptor-like. As mentioned in the introduction we do not differ here between defects at the interface and defects in the dielectric as we assume that the electrical response based on these defects may appear similar. It seems quite intuitive that the broad distribution of capture and emission times seen from MSM characterization suggests that the drift effects are related to defect states in the dielectric. While this is certainly a possibility, it may as well be that these broadly arranged defect states are within a thin interfacial layer at the III-N surface, as argued earlier. It seems currently that we have no evidence to claim either one in particular.

Based on the argument that the donor states are a separate type of defect without major contribution to the drift seen in PBTI experiments, we need to consider the unknown dynamic behavior of those native donor states. From the total donor state density ND, only the amount of ionized donors ND+ is known from the charge neutrality condition (Eq. (7)):

$$N_{D+}(E, x) = \frac{N_{D,0}}{\sigma_{B} \tau_{V}} \frac{e^{-\frac{E - \mu_{D}(x)}{2kT}}}{2\pi kT}$$

where $N_{D,0}$ is the concentration of occupied donors below the Fermi level. As discussed earlier in the Introduction, it is commonly believed that those defects exist. From the fluorinated donor state density, we derived the total amount of ND to be $3.3 \times 10^{13} \text{ cm}^{-2}$. Interestingly this number is close to the polarization charge ($2.9 \times 10^{13} \text{ cm}^{-2}$) of the AlGaN barrier. However, it would be surprising if a link would exist between the polarization charge and the surface donor density. Such an effect could only be explained if the donor states are formed during the epitaxial process and if their formation would be linked to the polarization charges at the barrier/channel interface. However, the contribution in lowering the formation energy per surface atom would be in the order of 1/100 of the energy of an electron. Therefore it is rather unlikely that the polarization charge directly influences the formation of the donor states. As it is very unlikely that the amount of donor states...
is exactly same as the required number of ionized donors to create the 2DEG, we argue that $N_{D0} > 0$, most likely in the range of low $10^{13} \text{cm}^{-2}$ and therefore should be considered for the device behavior. This leads then to two new questions: Why do the ionized donors $N_{D0}$ not capture an electron during forward gate bias? And why do the occupied donor states (in standard devices) not emit their electron during negative gate bias? These questions are particularly interesting as NBTI is not only relevant for the gate in MIS devices but also in the access region of lateral GaN-based devices relying on a 2DEG. Under negative bias or off-state conditions those areas between gate and drain would be depleted and therefore experience similar vertical bias conditions as the dielectric interface below the gate. Most reports discussing NBTI cannot be exclusively linked to the interface and discuss combined charge trapping effects with defects in the buffer [142,143]. Considering the defect model in Fig. 1b, we can explain the lack of contribution of the native donor states in NBTI and PBTI processes in standard MISHEMT by sufficiently high defect barrier energies for electron capture $\Delta E_{c}$ and emission $\Delta E_{em,apparent}$ (Fig. 37). Considering the electron emission from the SRH model, a typical defect cross-section range from $10^{-12} \text{ cm}^{-2}$ and $10^{-18} \text{ cm}^{-2}$, and lifetime conditions for 10 years at 150 °C, we would require an energy barrier of 2.1 to 2.6 eV (Fig. 38). As there are very few pieces of information available to model the donor states at the III-N surface by ab-initio simulation methods, we believe this is an important requirement to be taking into account. Following our arguments about the native donor states being not involved in PBTI or NBTI effect, we should note that the commonly observed charge trapping effects at the interface with respect to current collapse or dynamic $R_{on}$ are likely linked to the ones seen in PBTI effects.

6. Conclusions

We have discussed that a lot of accuracy and precision is needed to characterize defects at the III-N interfaces. MSM is clearly the best choice to compare and characterize the typically broad distribution of time constants for capture and emission. 2nd-order kinetics effects can be recognized due to multi-state defects as well as the Coulomb effect modifying the interface potential during trapping and hence the effective active energy area. Furthermore, the effective capture and emission time can be influenced by the electron transport above the III-N barrier. A worst-case life time criterion was introduced considering the effect of the 2DEG depletion below the gate which increases the on-resistance and decreases the maximum drain-source current. Under this criterion, the maximum interface density for end-of-life conditions has been estimated to about $4-8 \times 10^{13} \text{cm}^{-2}$, mostly independent of the dielectric material. Nevertheless, today there seems no proven solution that would fulfill this lifetime criterion. Fluorinated surface modification has resulted in drastically reduced capture and emission times, caused by an alteration of the native surface donors. This highlights the still open question on how native donor states maintain unionized/occupied over lifetime at the dielectric interface below the gate or in the access region between gate and drain.

Acknowledgement

We would like to acknowledge the support of Gerhard Prechtl, Michael Stadtmüller, Günter Denifl, Helmut Angerer, and Oliver Häberlen from Infineon Technologies Austria AG, Gregor Pobegen, and Roberta Stradiotto from KAI, Alexander Grill, Mattia Capriotti, and Michael Schmid from Vienna University of Technology, Rudolf Pletschig from University of Kassel, Erhard Kohn from University of Ulm, Liverios Lyperakis and Jörg Neugebauer from Max-Planck-Institut (MPIE) in Düsseldorf. Further we would like to thank the editors Tibor Grasser and James H. Statth as well as the reviewers for their valuable inputs.

References


Fig. 38. Calculated emission time of donor state defects for different ionization energies at 300 and 450 K and capture cross-sections of $10^{-14} \text{cm}^2$ and $10^{-18} \text{cm}^2$. 

80
C. Ostermaier et al.


