A Rigorous Study of Measurement Techniques for Negative Bias Temperature Instability

Tibor Grasser, *Senior Member, IEEE*, Paul-Jürgen Wagner, Philipp Hehenberger, Wolfgang Goes, and Ben Kaczer

***Abstract*—The active research conducted in the last couple of years demonstrates that negative bias temperature instability is one of the most serious reliability concerns for highly scaled pMOSFETs. As a fundamental prerequisite for a proper understanding of the phenomenon, accurate measurements are indispensable. Unfortunately, due to the nearly instantaneous relaxation of the degradation once the stressing conditions are removed, an accurate assessment of the real degradation is still extremely challenging. Consequently, rather than interrupting the stress in order to measure the degradation, alternative measurement techniques, such as the on-the-fly methods, which avoid stress interruption, have been proposed. However, these methods rely on rather simple compact models to translate the observed change in the linear drain current to a threshold voltage shift. As such, all methods have their own drawbacks which are rigorously assessed using a theoretical description of the problem.**

***Index Terms*—Interface states, measurement, measure/stress/measure (MSM), negative bias temperature instability (NBTI), on-the-fly (OTF), oxide charges.**

I. Introduction

WHEN SUBJECTED to negative bias temperature stress, device parameters of pMOS transistors have been observed to degrade [1]–[3]. This degradation is commonly described by a shift of the drain current or threshold voltage as a function of stress time. Due to the extremely fast recovery which sets in as soon as the stressing conditions are removed, any accurate assessment of the real degradation is very challenging.

In order to cope with that serious problem, many different measurement techniques have been proposed so far, which fall into different categories.

1) Conventional measure/stress/measure (MSM) techniques have been optimized by minimizing the delay between the removal of the stress and the first measurement point. In MSM techniques, either the drain current around the threshold voltage [4] is recorded and mapped to $\Delta V_{\rm th}$ using an initial $I_{\rm D}-V_{\rm G}$ curve around $\Delta V_{\rm th}$, or a threshold current is enforced through the device [5], thereby directly measuring $V_{\rm th}$. Alternatively, complete $I_{\rm D}-V_{\rm G}$ curves have been recorded using ultrashort pulses [6], [7]. It is still unclear whether it is possible to reduce the delay sufficiently in order to record the true degradation characteristics [8], [9].
2) On-the-fly (OTF) techniques monitor the degradation of the drain current in the linear regime $I_{\rm Dlin}$ directly under stress conditions, thereby avoiding any relaxation. Different variants have been put forward [2], [10]–[12] which are more or less straightforward to implement [12].
3) In order to differentiate between oxide and interface charges, specialized measurement techniques such as charge-pumping (CP) and DCIV techniques have been used [3], [13]. However, due to the unclear interference of positive bias with the recovery characteristics, it is still very difficult to correctly interpret such data.

In order to link the observed degradation to possible underlying physical mechanisms, most commonly the creation of interface states and positive oxide charges [2], [3], a parametric relationship between the threshold-voltage shift/drain current and these charges is required. The natural candidate is the shift of the threshold voltage in the subthreshold regime, because it is directly related to the effective density of interface states and oxide charges via Poisson's equation as

$$\Delta V_{\rm th}(t) = -\frac{\Delta Q_{\rm it}(t) + \Delta Q_{\rm ot}(t)}{C_{\rm ox}}. \tag{1}$$

The positive charge stored in the interface states $Q_{\rm it}(t)$ is given by $\mathrm{q}N_{\rm it}(t)f(V_{\rm G})$, where f denotes the hole occupancy of $N_{\rm it}$ depending on the gate (measurement) voltage $V_{\rm G}$.

By contrast, $I_{\rm Dlin}$ as recorded during OTF techniques not only depends on $Q_{\rm it}$ and $Q_{\rm ot}$, but also on the effective mobility $\mu_{\rm eff}$, which is known to change during degradation [1], [11], [14]. The three OTF extrapolation methods investigated here are based on a simple compact model (SPICE level-1)

$$I_{\rm D} = \frac{\beta V_{\rm D}\left(V_{\rm G} - V_\theta - \frac{1}{2}V_{\rm D}\right)}{1 + \theta\left(V_{\rm G} - V_\theta - \frac{1}{2}V_{\rm D}\right)}, \qquad \text{for } V_{\rm G} > V_\theta \tag{2}$$

or variants thereof, valid only in the linear regime under strong inversion [15].[1] The parameter β depends on the effective mobility, and θ describes the deviation from the linear behavior. Curiously, the fact that the threshold voltage V_θ used in expressions like (2) is the purely empirical *extrapolated threshold voltage* [12], [15] rather than the threshold voltage one obtains from the solution of the Poisson equation is rarely appreciated. This implies that changes in the extrapolated threshold voltage are not necessarily related in a 1:1 manner to additional charges

Manuscript received December 29, 2007; revised February 29, 2008. Current version published October 16, 2008.

T. Grasser, P.-J. Wagner, P. Hehenberger, and W. Goes are with the Christian Doppler Laboratory for TCAD in Microelectronics, Institute for Microelectronics, Technische Universität Wien, 1040 Wien, Austria.

B. Kaczer is with the Reliability Group, IMEC, 3001 Leuven, Belgium.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TDMR.2008.2002353

[1]We use the conventional nMOS sign convention in all formulas, with the sign change for pMOS devices being understood.

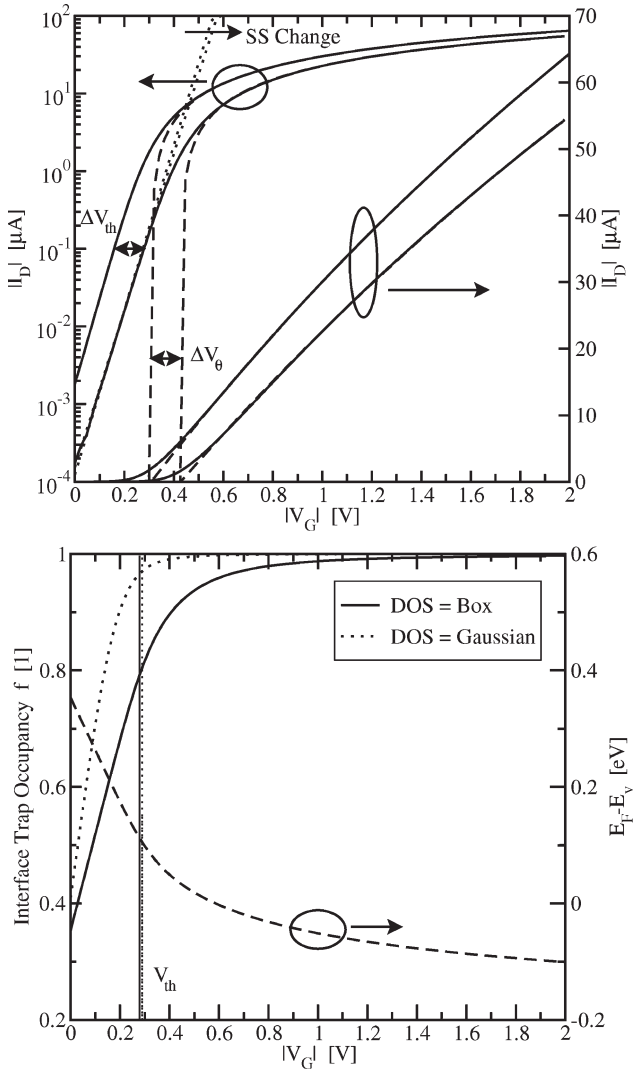

Fig. 1. (**Top**) Simulated $I_{\rm D}-V_{\rm G}$ characteristics before and after 10^5-s stress (shift to the right). It is assumed that, during stress, oxide and interface charges are accumulated ($N_{\rm ot}=1.3\times10^{12}$ cm^{-2}, $N_{\rm it}=7.2\times10^{11}$ cm^{-2}) which also lead to a 10% mobility degradation. Such an excessive degradation is, of course, unrealistic but results in a clearly visible impact and has been chosen for illustrational purposes only. It is clearly visible that the extracted ΔV_θ and $\Delta V_{\rm th}$ are different quantities, with V_θ being larger than $V_{\rm th}$ [15]. Also indicated is the small change in the subthreshold slope caused by the changing interface state occupancy which affects the extraction of $\Delta V_{\rm th}$. (**Bottom**) Interface state occupancy f as a function of the gate voltage. During the measurement around the threshold voltage (indicated by thin straight lines), most traps are occupied (80%) for a flat DOS in the lower half of the bandgap. Also, shown is the gate voltage dependence of a Gaussian DOS with a maximum around 0.235 eV above the valence band edge [17], [18]. For a Gaussian DOS, a sharper transition, which occurs even later and gives an occupancy of about 97%, is observed. In order to consider the worst case, a (box) flat DOS is assumed in the following. Finally, the difference of the Fermi level $E_{\rm F}$ from the valence band edge is shown, resulting from Boltzmann statistic, which overestimate the hole concentration in strong inversion.

at the interface/oxide, also due to hidden correlations in the parameters V_θ, θ, and β [16].

In order to emphasize this issue, we use different symbols for $V_{\rm th}$ and V_θ. The difference of these two quantities is shown in Fig. 1. In particular, V_θ is always larger than $V_{\rm th}$ [15]. Furthermore, the exact value of V_θ obtainable from a calibration depends on the truncation of the measurement data used, because (2) is only valid for larger $V_{\rm G}$. For instance, in [19], only data with $V_{\rm G}$ larger than the value of $V_{\rm G}$ where the maximum slope is obtained is used, in order to guarantee a good fit of (2) in the strong inversion region. We remark that excellent fits of (2) are indeed possible in the strong inversion region [12], with increasing inaccuracies when $V_{\rm G}$ approaches V_θ.

One word of caution is in order: In the following study, we consider the threshold voltage shift in the subthreshold regime $\Delta V_{\rm th}$ to be superior to ΔV_θ simply because it can be used to identify the buildup of interface and oxide charges via (1), provided possible distortions are small, because these charges are required for a correct physical understanding and modeling. Although from a circuit-designer point of view the shift in V_θ may be more relevant, it is more difficult to build a physics-based model on that data, as will be demonstrated in the following.

It is thus important to realize that none of the available measurement techniques is able to directly measure the real degradation; whereas MSM techniques have to cope with unavoidable delays, OTF techniques may also suffer from the conversion of $\Delta I_{\rm Dlin}$ to $\Delta V_{\rm th}$, which is conventionally justified [2], [11] using rather simple compact models. As such, the results given by these techniques have been reported to differ by more than an order of magnitude in extreme cases [3], [11]. Although this is commonly attributed to the delay inherent in MSM techniques [3], we will show that this issue is more involved by conducting a thorough theoretical study using a numerical device simulator enhanced by appropriate reliability models.

II. Physical Modeling

A comparison of MSM and OTF results is shown in Fig. 2 for a typical IMEC device [4]: While the MSM technique can estimate the true degradation only via extrapolation (we use our technique [9], [20] based on the universality of negative bias temperature instability (NBTI) degradation [20], [21]), the $\Delta V_{\rm th}$ values predicted by the simplest OTF method do not fit the detailed relaxation data very well, regardless of any extrapolation uncertainties.

In order to study this discrepancy, we turn the problem around: We place a certain predefined amount of $\Delta Q_{\rm it}$ and $\Delta Q_{\rm ot}$ at the interface of a pMOS and numerically simulate the drain current observed in the various techniques. For our simulations, we employ a suitably extended version of our numerical device simulator MINIMOS-NT [22]. Carrier transport is described using the drift-diffusion formalism which delivers accurate drain currents for small $V_{\rm D}$ and not too small channel lengths [23]. Boltzmann statistics are assumed for the carrier concentrations, resulting in an overestimation for large gate overdrive. The impact of these assumptions to the assessment of OTF data needs to be considered in greater depth, although we do not expect the inclusion of Fermi–Dirac statistics to significantly alter our results. We include accurate models for the interface state dynamics (SRH) with a realistic density-of-state (DOS) distribution [20], [24] and the mobility variation due to Coulomb scattering induced by interface states [25]. For pure silicon dioxide, interface states have often been assumed to introduce donorlike Gaussian peaks close to the valence band edge (we use 0.235 eV [17], [18]) into the DOS. For the device

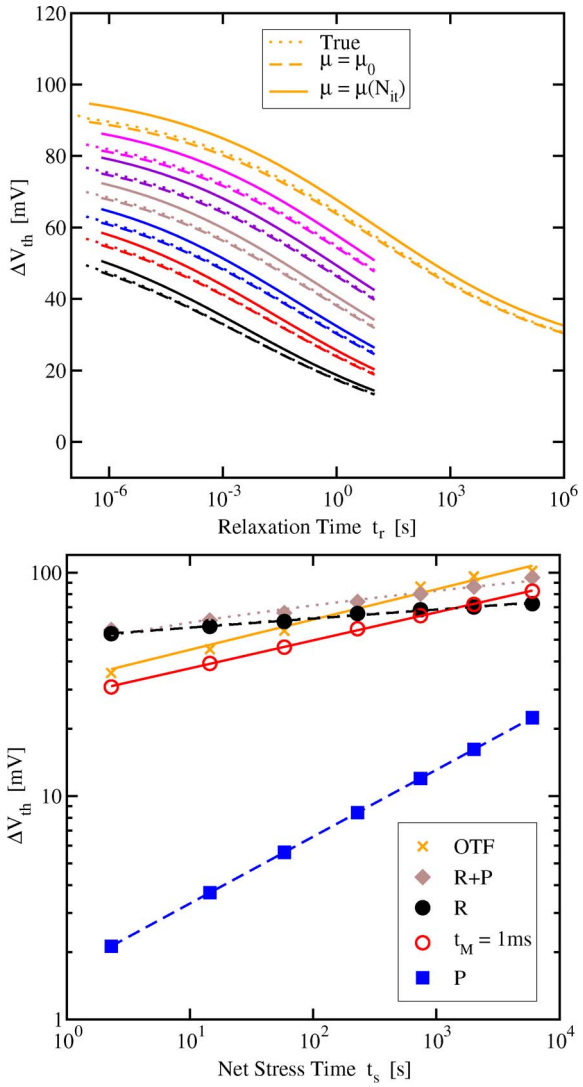

Fig. 2. Comparison of (open symbols) measured MSM and OTF1 data. (**Top**) Recorded detailed relaxation sequence allows one to reconstruct the (closed symbols and lines) full NBTI degradation by identifying a universally relaxing component R which is observed on top of a permanent/slowly relaxing component P. The degradation observed in OTF1 measurements does not agree well with the relaxation data. (**Bottom**) Extracted components R and P, the OTF1 data, and the MSM data with a delay of $t_M = 1$ ms as a function of the stress time. Clearly, the overall degradation $R + P$ is only in poor agreement with the OTF1 data.

under consideration, this implies that, around the threshold voltage, practically all interface states are occupied by holes. Recently, it has been pointed out that the DOS in oxynitrides may be markedly different [24]. Because the occupancy of the interface states determines the "visible" charge in the interface states, a changing occupancy during $I_{\rm D}-V_{\rm G}$ measurements changes the subthreshold slope and may consequently distort the extracted $V_{\rm th}$. In order to consider the worst case, we use a flat DOS in the lower half of the bandgap in the following, the impact of which is compared to a Gaussian distribution in Fig. 1. Furthermore, all simulations are based on a numerical device description carefully calibrated to measured $I_{\rm D}-V_{\rm G}$ curves for a pMOS with $L_g = 0.5$ μm and EOT = 1.4 nm.

The simulated drain current is then processed analogously to the measurement data and converted into $\Delta V_{\rm th}$. We then compare whether the true $\Delta V_{\rm th}$ given by the predefined charges agrees with what is observed in the simulated measurements. Consequently, this approach allows us to study the sensitivity of the extracted $\Delta V_{\rm th}$ to the measurement setup and possible distortions.

Lacking a valid NBTI model that covers the degradation during both stress and recovery, as needed for simulated measurements, we rely on our empirical model in the following [9]: A growing number of recent publications suggests [2], [3], [26] that the degradation caused by NBTI consists of two components, a fast universally recoverable component R on top of a slowly recovering or permanent component P [8], [9]. The mapping of these two components on oxide or interface charges is still under debate, but it is often suggested that the recoverable component R is due to oxide charging and discharging, whereas P is caused by interface states [2]. Alternatively, it has been suggested that NBTI is primarily determined by interface states [3].

We empirically describe these two components as

$$R(t_s, t_r) = A \log\left(1 + \frac{t_s}{\tau_R}\right) r(t_r/t_s) \tag{3}$$

$$P(t_s) = \frac{P_{\max}}{1 + (t_s/\tau_P)^{-\alpha}} \tag{4}$$

with the universal relaxation function

$$r(\xi) = \frac{1}{1 + B\xi^{\beta_r}} \tag{5}$$

and the parameter values $A \approx 2.7$ mV, $\tau_R \approx 1.4$ ns, $B \approx 3.2$, $\beta_r \approx 0.16$, $P_{\max} \approx 50$ mV, $\tau_P \approx 13$ ks, and $\alpha \approx 0.45$. We remark that these values are determined for relaxation at $V_{\rm G} \approx V_{\rm th}$ and vary for other bias conditions. For the extraction of the parameters, the algorithm detailed in [8] and [9] has been applied to the data shown in Fig. 2. Consequently, our empirical description is accurate only under this condition and does not describe relaxation at varying $V_{\rm G}$ (as would be needed for a correct analysis of $I_{\rm D}-V_{\rm G}$ and CP data). In the following, we will consider the two cases:

1) The degradation is due to oxide and interface states, assuming $\Delta Q_{\rm ot}(t_s, t_r) = R(t_s, t_r)$ and $\Delta N_{\rm it}(t_s) = P(t_s)$.
2) The degradation is due to interface states only, $\Delta N_{\rm it} = R(t_s, t_r) + P(t_s)$.

The bias dependence of the oxide charges is hardwired in the universal relaxation function for a recovery around $V_{\rm G} \approx V_{\rm th}$, whereas the actual charge stored in the interface states depends on the position of the Fermi level (and thus the gate voltage) via the density-of-states. For our choice of parameters [24], we found that during stress all interface states are positively charged, whereas during relaxation (measurement around $V_{\rm th}$), the occupancy goes down to ~80% (see Fig. 1).

For the sake of completeness, we remark that the widely used reaction-diffusion model [1], [3], given by $R = At_s^n r(t_r/t_s)$, $n = 1/6$, $B = 1$, and $\beta_r = 1/2$, is unable to capture any relaxation data known to us [9], [20].

In the following, the measurement methods discussed in this paper are summarized. We also highlight potential distortions due to often unappreciated physical effects like mobility variations and interface state dynamics.

A. MSM Methods

First, we use the method suggested in [4] which records several relaxation phases after exponentially growing stress intervals. As such, a maximum amount of information on a single device in a single measurement cycle is gathered. The recorded drain current around $V_{\rm th}$ is then converted into $\Delta V_{\rm th}$ using an initial (narrow) $I_{\rm D}-V_{\rm G}$ curve. Using the method developed in [8], a fast component R on top of a permanent component P is extracted and mapped to $\Delta Q_{\rm ot}$ and $\Delta Q_{\rm it}$ using the previously stated prescriptions.

In addition to this $I_{\rm D}@V_{\rm th}$ method, we calculate fast ($1\ \mu$s) $I_{\rm D}-V_{\rm G}$ curves for the extraction of $\Delta V_{\rm th}$ using two constant current criteria. We remark that the relaxation predicted by our model is accurate only when switching the bias from the stress level to $V_{\rm G}=V_{\rm th}$, and the estimates obtained using this method must be taken with care.

B. OTF Methods

The first and simplest OTF method we discuss (OTF1) assumes θ and $V_{\rm D}$ to be small and also neglects any change in the effective mobility $\mu_{\rm eff}$ (and thus β) [1], [3]

$$\Delta V_\theta^{(1)} \approx \frac{I_{\rm D}-I_{\rm D0}}{I_{\rm D0}}(V_{\rm G}-V_{\theta 0}). \tag{6}$$

The OTF1 method has the advantage that it only requires the determination of the initial drain current in the linear regime under stress conditions. The neglect of the mobility change (β and θ) in the OTF1 method has motivated the proposal of the OTF2 method which is capable of monitoring a mobility variation by introducing small perturbations of the stress voltage [2]

$$\Delta V_\theta^{(2)} \approx \sum_n \frac{\frac{I_{\rm D}[n]}{g_{\rm m}[n]}-\frac{I_{\rm D}[n-1]}{g_{\rm m}[n-1]}}{1+2\theta\left(V_{\rm G}-V_\theta[n]-\frac{1}{2}V_{\rm D}\right)}. \tag{7}$$

However, OTF2 still neglects the temporal change of θ. Furthermore, the OTF2 technique requires the determination of a full initial $I_{\rm D}-V_{\rm G}$ curve for the determination of θ. Thus, as an alternative, one may use a variant suggested by Zhang and Chang [11] (OTF3)

$$\Delta V_\theta^{(3)} \approx \sum_n \frac{I_{\rm D}[n]-I_{\rm D}[n-1]}{\frac{1}{2}\left(g_{\rm m}[n]+g_{\rm m}[n-1]\right)}. \tag{8}$$

III. Possible Distortions

The analysis of measurement data often assumes that the change in the drain current is only caused by ΔV_θ due to oxide/interface charges. However, it has been pointed out [2], [11], [14], [27] that mobility variations may impact the accuracy of the methods. This issue is commonly associated with OTF techniques, but a rigorous theoretical analysis has not been conducted so far. This will be done in the following, where we also demonstrate the impact of mobility variations on MSM techniques. In addition, the response of the interface states on rapidly changing gate voltages is discussed.

A. Mobility Variations

1) OTF Methods: To analyze the sensitivity of the extracted ΔV_θ to changes in the mobility, we estimate the drain current in the linear region by a further simplified version of (2) ($\theta\approx 0, V_{\rm D}\ll V_{\rm G}$)

$$I_{\rm D}=\beta V_{\rm D}(V_{\rm G}-V_\theta). \tag{9}$$

According to OTF1, the change in the linear drain current is converted to a threshold-voltage shift using $\Delta V_\theta=\Delta I_{\rm D}/\beta$. Because β is approximately linearly proportional to $\mu_{\rm eff}$, according to (9), a change in $\mu_{\rm eff}=\mu_{\rm eff0}+\Delta\mu_{\rm eff}$ results in

$$\Delta I_{\rm D}=I_{\rm D0}\Delta\mu_{\rm eff}/\mu_{\rm eff0}. \tag{10}$$

Assuming that this shift $\Delta I_{\rm D}$ is *only* caused by mobility variations, a spurious shift in ΔV_θ is obtained as

$$\Delta V_\theta^\mu=-\frac{\Delta\mu_{\rm eff}}{\mu_{\rm eff0}}(V_{\rm G}-V_\theta)=-(V_{\rm G}-V_\theta)\frac{p}{100} \tag{11}$$

with p being the mobility variation in percent. With $V_{\rm G}=2$ V and $V_\theta\approx 0.25$ V (fit to the linear part of the on-current), we obtain that the error in the extracted ΔV_θ equals 17.5 mV for each percent mobility error, rather much larger than previously anticipated [1]. As an example, for an error in the effective mobility of 3% [1], one obtains an error of about 50 mV in ΔV_θ, which is considerable keeping in mind that a 100 mV shift is often considered to cause failure.

2) MSM Methods: By contrast, data obtained via the MSM technique is much less sensitive to mobility changes. This can be easily demonstrated starting from a simple compact model for the subthreshold current [15]

$$I_{\rm D}=A\mu_{\rm eff}\exp\left(\frac{V_{\rm G}-V_{\rm th}}{\eta V_T}\right). \tag{12}$$

Here, V_T is the thermal voltage $k_BT/{\rm q}$. Although the current in the subthreshold regime is also proportional to the mobility change and, consequently, a change in the mobility results also in an equivalent change of $I_{\rm D}$ via (10), the drain current depends now *exponentially* on the threshold-voltage shift. Thus, by inverting (12), one obtains

$$\Delta V_{\rm th}^\mu=-\eta V_T\log\left(1+\frac{\Delta\mu_{\rm eff}}{\mu_{\rm eff0}}\right)\approx-\eta V_T\frac{\Delta\mu_{\rm eff}}{\mu_{\rm eff0}} \tag{13}$$

$$=-\eta V_T\frac{p}{100}. \tag{14}$$

For the device under consideration, $\eta\approx 1.14$ (subthreshold slope of about 68 mV/dec at room temperature), giving $\eta V_T\approx$ 42 mV and, finally, a 420-μV error for each percent variation in the mobility. However, because the variation of the low-field mobility is expected to be larger than the high-field mobility [1], [11], for instance, 10%, one obtains an error in $\Delta V_{\rm th}$ of about

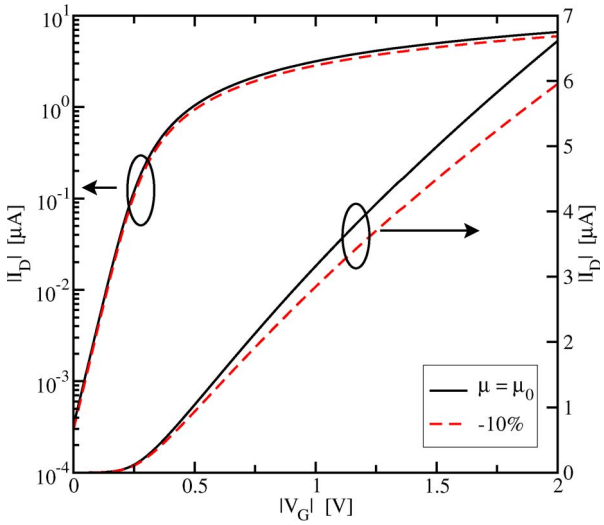

Fig. 3. Simulated $I_{\rm D}-V_{\rm G}$ characteristics at $V_{\rm D}=-50$ mV with nominal mobility values and an excessively large mobility degradation of about 10% for illustrative purposes. Clearly, the impact is largest in the linear regime.

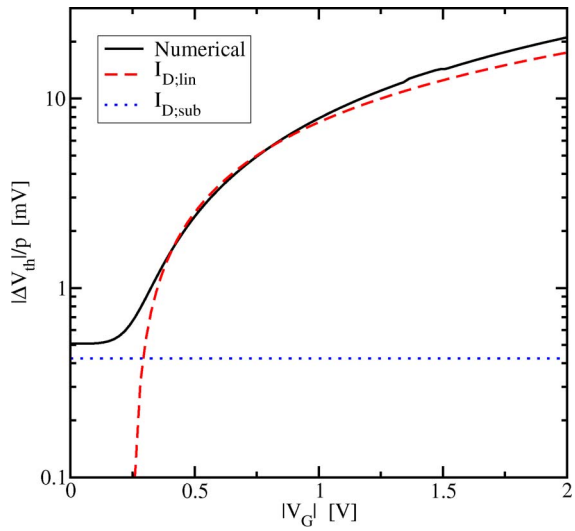

Fig. 4. Spurious threshold-voltage shift ΔV_θ^μ per percent change in $\mu_{\rm eff}$ obtained from Fig. 3 by considering the difference in $V_{\rm G}$ at the same $I_{\rm D}$. With increasing $V_{\rm G}$, ΔV_θ^μ increases. Also, shown are the error estimates for the (dashed lines) linear regime (11) and the (dotted lines) subthreshold regime (14).

5 mV, which is about a factor of ten lower than the corresponding error in the OTF method. Interestingly, the mobility-variation-induced error of the MSM method depends linearly on temperature via V_T, whereas the error of the OTF has only a weak temperature dependence via $V_{\rm th}$.

3) General Gate-Voltage Dependence for Mobility Variations: The previous two results clearly show that the impact of the mobility variation on the estimated threshold-voltage shift depends on the operation mode of the transistor and, thus, on the applied gate voltage. Fig. 3 shows the change in the numerically simulated $I_{\rm D}-V_{\rm G}$ characteristics caused by a 10% mobility degradation. Clearly, the impact is largest in the linear regime. The spurious threshold-voltage shift caused by this mobility variation is obtained by determining the difference in $V_{\rm G}$ at the same current $I_{\rm D}$. For reasonable changes in the mobility, this spurious $\Delta V_{\rm th}^\mu$ depends linearly on the mobility variation. Fig. 4 shows $\Delta V_{\rm th}^\mu$ per percent mobility variation p as a function of gate bias. The smallest dependence of $\approx 500\,\mu{\rm V}\times p$ is obtained in the subthreshold regime (in that example, $V_{\rm G}<0.2$ V), where $I_{\rm D}$ perfectly follows (12). This value is in good agreement with the analytic estimate given by (14). Consequently, in order to avoid mobility-induced distortions of $\Delta V_{\rm th}$, measurements should be conducted with $V_{\rm G}$ safely in the exponential regime of $I_{\rm D}$. For larger $V_{\rm G}$, on the other hand, $\Delta V_{\rm th}^\mu$ rapidly increases in accordance with (11), which is also in good agreement with the full numerical result.

B. Dynamic Response of the Interface State Occupancy

An issue particularly relevant for ultrafast MSM measurements is the response of the interface states to a quickly changing gate voltage. It has been argued that, for measurements in the microsecond regime, the interface states may not react fast enough to the rapid drop in $V_{\rm G}$ and thus distort the response of $I_{\rm D}$ and, consequently, the estimated $\Delta V_{\rm th}$ [27], [28]. From a theoretical point of view, the dynamics are governed by dynamic SRH kinetics, which is straightforward to include in a

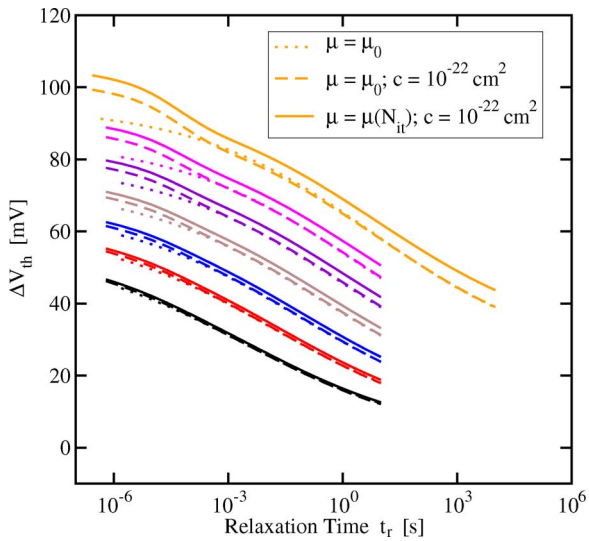

Fig. 5. Simulated impact of a 10% mobility variation and the dynamics of the interface state occupancy on the extracted $\Delta V_{\rm th}$. The same stress times as in Fig. 2 are used. Because the mobility model used depends only on $N_{\rm it}$, which is assumed to be permanent, the mobility variation causes a constant offset in each relaxation curve (4.5 mV in the last curve at $t_s=6000$ s). Also, only for unusually small capture cross sections c on the order of 10^{-22} cm^2, initial distortions of $\Delta V_{\rm th}$ can be observed.

numerical simulation. Again, a flat DOS is used for the interface states in order to maximize the impact of this effect (worst-case estimation). The timescales are determined by the capture cross sections, somewhat elusive quantities [29], and by how far the Fermi level is moved away from the valence band edge. In order to accurately reproduce CP measurements, capture cross sections on the order of 10^{-16} cm^2 are used in our simulator. Note, however, that values in the range $10^{-14},\ldots,10^{-18}$ cm^2 have been reported [29], [30]. Interestingly, an impact of interface state dynamics on the simulated $I_{\rm D}$ in the microsecond regime is only observed for unusually small capture cross sections on the order of 10^{-22} cm^2 (cf. Fig. 5), in disagreement with the

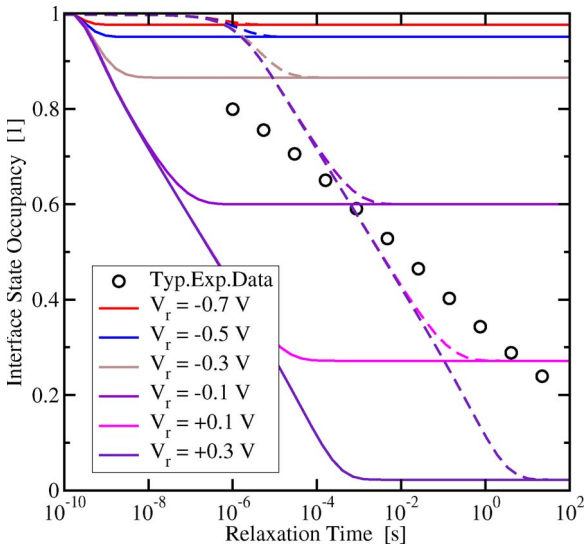

Fig. 6. Simulated response of the interface state occupancy when $V_{\rm G}$ is switched to a smaller recovery voltage. The dashed lines are obtained with extremely small capture cross sections of 10^{-22} cm^2.

assertions made in [27]. Such small capture cross sections can only be justified by assuming defects at a certain distance from the interface which require tunneling for charge exchange [31]. We also remark that such a process, if occurring on top of the fast component R, would result in a visible distortion of the relaxation of $I_{\rm D}$, not present in any of the data available to us [9]. Thus, we prefer the notion that if such a process is relevant at all, it should naturally be part of the fast component R.

To understand the dynamic response more clearly, it is intuitive to investigate the step response of the interface states when $V_{\rm G}$ is switched from stress bias (-2 V) to various recovery biases (cf. Fig. 6). Although the response is "loglike" [27], it is too fast to be significant for negative recovery voltages and conventionally used capture cross sections.

IV. Results

In the following, we introduce the analytic models for $Q_{\rm ot}$ and $Q_{\rm it}$ into the device simulator, resimulate the various measurement sequences, and compare the extracted with the expected values.

A. MSM Methods

We begin with the MSM method used to originally extract the temporal evolution of $Q_{\rm ot}$ and $Q_{\rm it}$. Quite convincingly, the simulation reproduces the relaxation data at $V_{\rm G} \approx V_{\rm th}$ and captures the trend in the OTF1 data at $V_{\rm G} = -2$ V (cf. Fig. 7). The extracted $\Delta V_{\rm th}$ with zero delay (frozen relaxation) and neglected mobility changes agrees perfectly with the expected values given through $Q_{\rm ot}$ and $Q_{\rm it}$. However, when the degradation of the mobility is considered (about 10% at 10^5 s [1]), an error of about 4 mV at $t_s = 6000$ s is observed. Because the mobility model used assumes only the permanent interface states to affect the mobility, no relaxation in this error is observed. Consequently, this error translates directly to the permanent component. In reality, oxide charges located farther

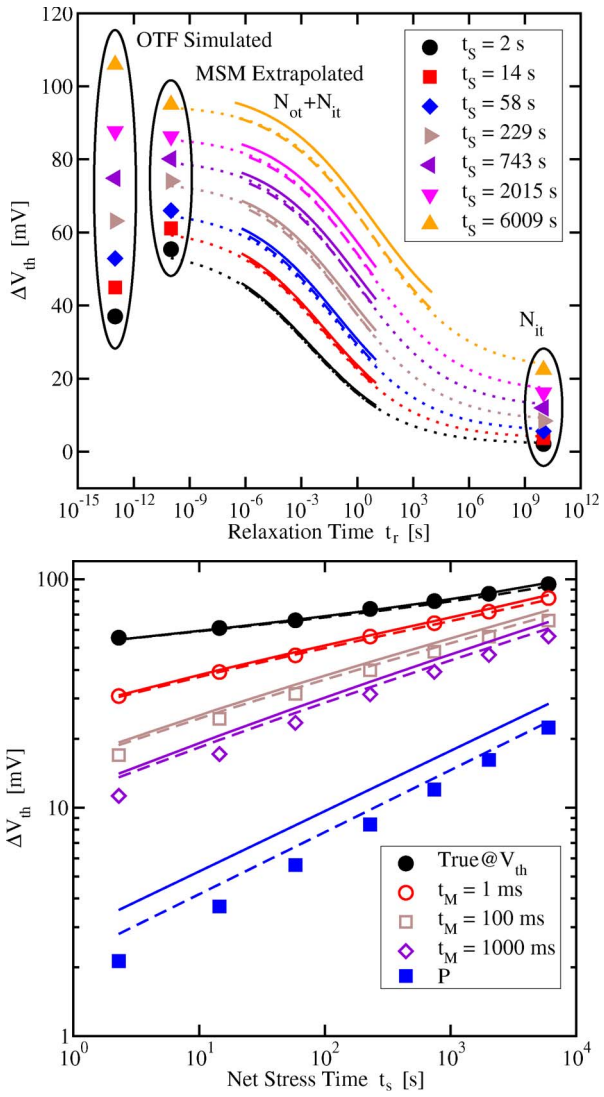

Fig. 7. Simulation of the measurement sequence shown in Fig. 2, which exactly reproduces the measurement results used to extract $N_{\rm it}$ and $N_{\rm ot}$ without any serious distortion. (**Top**) Extracted $\Delta V_{\rm th}$ shift, (solid and dashed lines) with and without considering the mobility variation, confirming that the extraction is only weakly dependent on the mobility variation. The dotted lines are a guide to the eye taken from Fig. 2. (**Bottom**) Same simulation results as a function of the stress time. The (lines) simulation results agree very well with the (open symbols) measurement data and the (closed symbols) extracted P and R components. A 4-mV error is introduced in the permanent component ($N_{\rm it}$) when the 10% mobility degradation is considered.

away from the interface also will affect the mobility; therefore, a slight change in this result is to be expected.

In addition to confirming the good accuracy of the MSM method, the influence of the measurement delay is exactly reproduced. In particular, although strongly dependent on the measurement delay, the extracted $\Delta V_{\rm th}$ is practically independent of the method used (conversion of $I_{\rm D}(V_{\rm th})$, see [4], various constant $I_{\rm D}$ criteria, ultrafast pulses, etc.) and the stress-dependent change in the mobility (cf. Fig. 8).

B. OTF Methods

A more involved behavior is observed for $\Delta V_{\rm th}$ extracted using the three OTF methods under consideration (cf. Fig. 9).

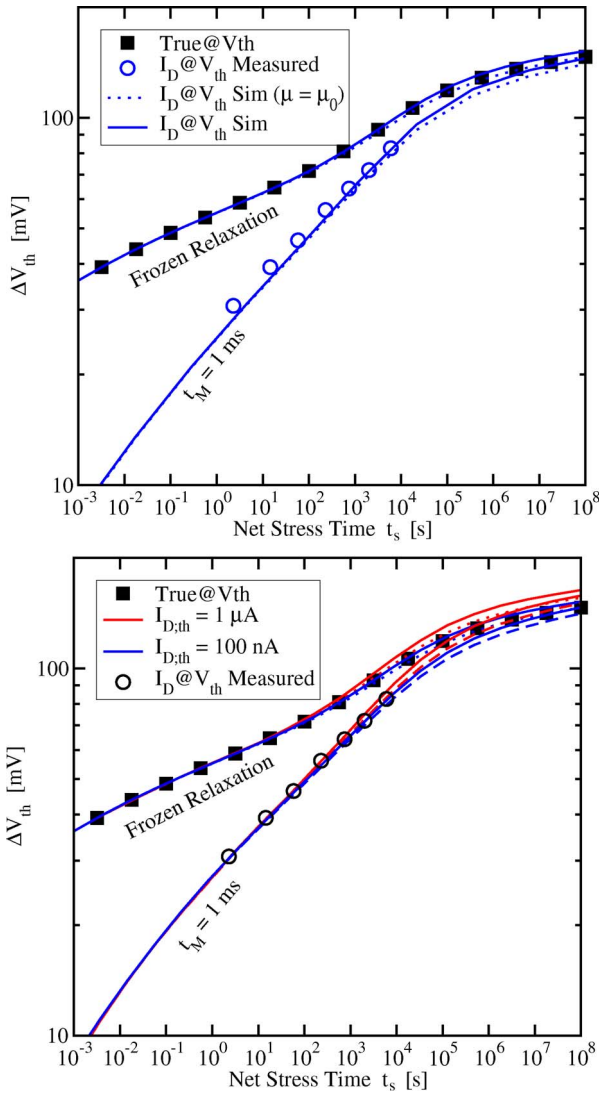

Fig. 8. Ideal no-delay MSM measurement exactly reproduces the true $\Delta V_{\rm th}$. In particular, (solid versus dotted lines) the extracted $\Delta V_{\rm th}$ is only weakly dependent on the mobility change and the extraction method. Shown are extractions (top) from $I_{\rm D}$ at $V_{\rm th}$ and (bottom) from $I_{\rm D}-V_{\rm G}$ curves with various $I_{D,\rm th}$ values. With increasing $I_{D,\rm th}$, a small additional error is visible due to the increased occupancy of $N_{\rm it}$.

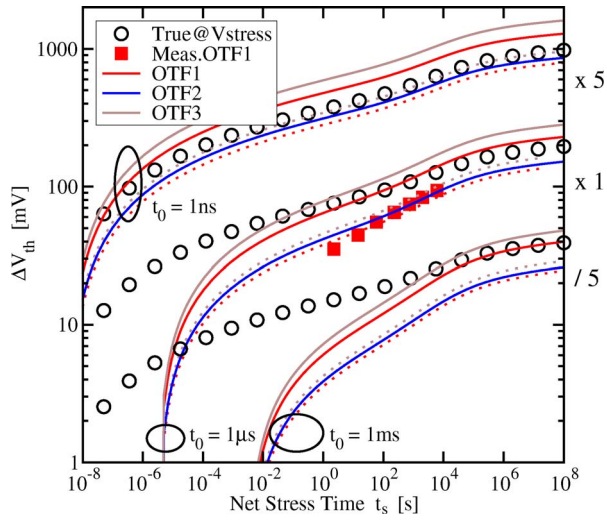

Fig. 9. Influence of the initial delay t_0 and the mobility change on OTF results. Whereas all OTF methods are strongly affected by the delay of the initial measurement (data scaled by a factor of five for readability), only OTF2 is not affected when (solid lines) the full interface charge dependent mobility model is used.

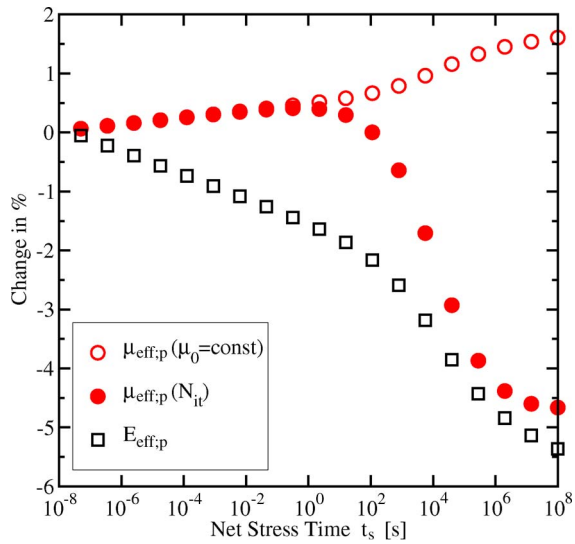

Fig. 10. Time dependence of the effective mobility ($\int \mu_p p{\rm d}y/\int p{\rm d}y$) and the effective electric field ($\int E_y p{\rm d}y/\int p{\rm d}y$) during the simulated measurement. Here, p is the hole concentration in the channel. When the influence of the interface charges is turned off in the mobility model, the effective mobility *increases* due to the reduction of $|V_{\rm G}-V_{\rm th}|$, (see [11]), whereas the inclusion of the additional Coulomb scattering at the interface leads to a degradation of about 4% at $t_s = 10^6$ s (see [1]).

All OTF methods suffer from the delay inherent in the first measurement point [32] which strongly distorts the data and is the equivalent problem of the delay in MSM techniques. We remark, however, that an exact prediction of the influence of the initial delay is difficult and may be exaggerated by assuming the universal relaxation law to be valid also for very short times.

In addition, the three techniques produce markedly different results, with only the OTF2 method [2] being insensitive to mobility changes (cf. Fig. 10). Note also that even extremely fast determination of the initial value $I_{\rm D0}$ leaves a residual error in the OTF methods due to the use of compact models. This is a rather severe disadvantage considering the widespread use of the OTF1 method in particular [1], [3].

The deficiencies of the OTF methods are also shown in Fig. 11 where the extracted slopes for OTF1 and OTF3 show a considerable error due to the mobility change, in contrast to the MSM method. In addition, we wish to remark that the theoretical accuracy obtainable by the OTF2 method is difficult to realize in practice, because the extracted $\Delta V_{\rm th}$ is extremely sensitive to unavoidable measurement errors in the estimation of $g_{\rm m}$ [12].

Finally, we check the potential errors when R and P are assumed to be due to interface states only. Fig. 12 shows the impact of a 10% mobility variation on the extracted $\Delta V_{\rm th}$. In contrast to the first case where oxide and interface charges where assumed, the mobility degradation causes a constant 5% error in the extracted data. Similarly, for the OTF methods,

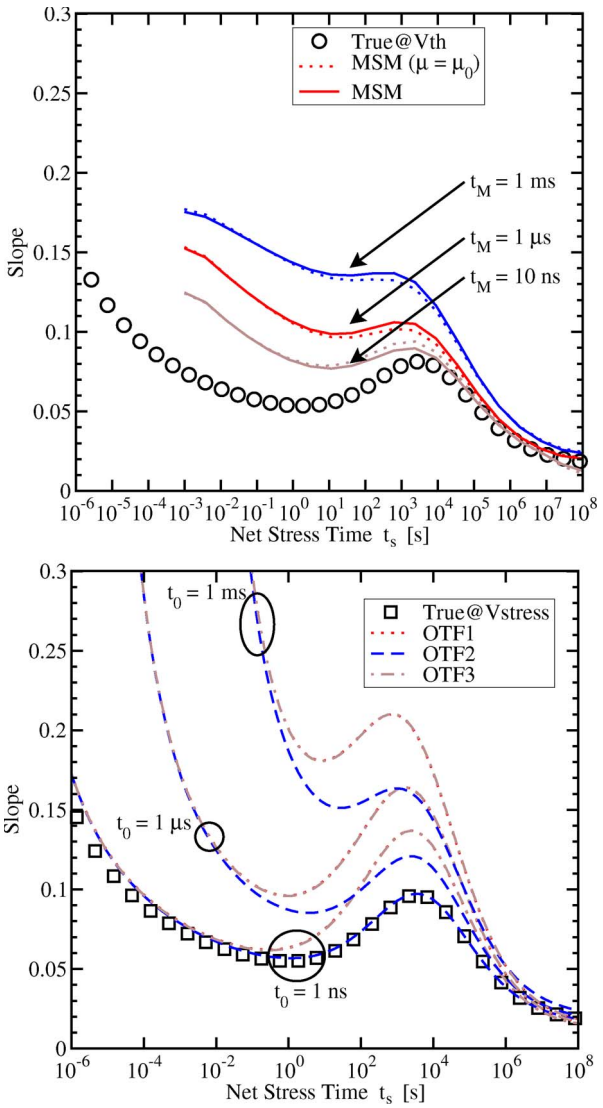

Fig. 11. Influence of the measurement method on the observed slope of an effective power-law time dependence. Whereas the (top) MSM methods depend primarily on the measurement delay t_{M}, OTF1 and OTF3 are contaminated by t_0, the mobility variation, and compact modeling errors ($\theta = \text{const}$, etc.). Only an extremely fast OTF2 method with a hypothetical initial delay of 1 ns is able to reproduce the expected values. The fastest hypothetical MSM method used a measurement delay of 10 ns; smaller delays resulted in distortions due to interface state dynamics.

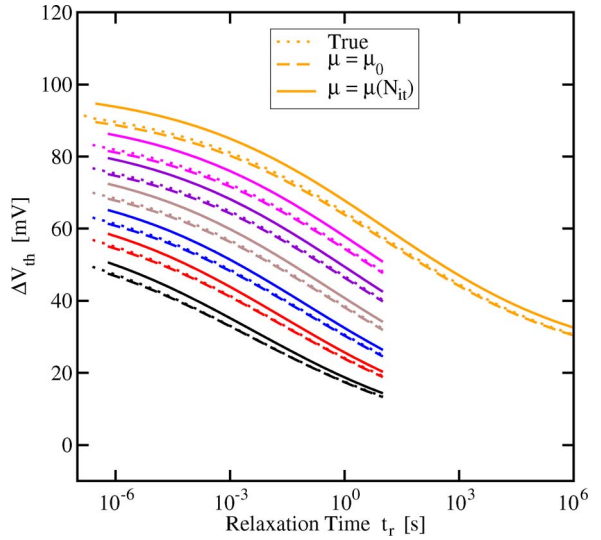

Fig. 12. The same as Fig. 5, but assuming that only interface states determine the degradation, $N_{\mathrm{it}} = R + P$. Now, the 10% mobility variation introduces a constant 5% error, rather than a 4.5-mV offset.

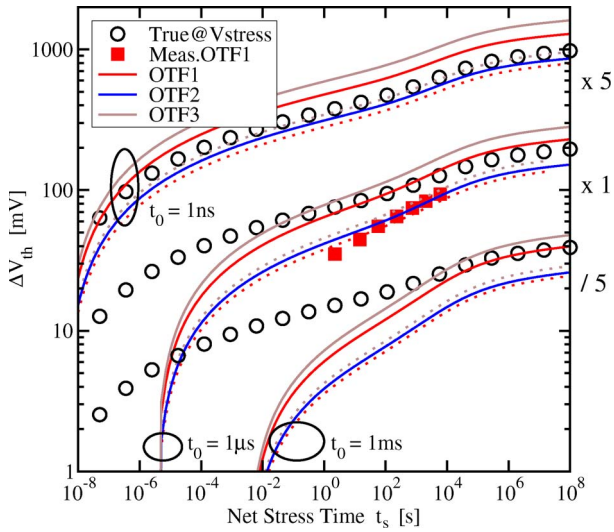

Fig. 13. The same as Fig. 9, but assuming that only interface states determine the degradation, $N_{\mathrm{it}} = R + P$. Because the mobility model assumes a mobility degradation as a function of N_{it}, a distortion is also obtained at earlier times.

Fig. 13 shows that, due to the early increase in ΔN_{it}, a distortion in the extracted ΔV_θ is obtained also at earlier times.

V. Conclusion

We have analyzed the most commonly used NBTI measurement techniques. Our analysis is based on a rigorous numerical solution of the semiconductor device equations, augmented by suitable models for interface and oxide charges. As expected, it is confirmed that MSM methods are distorted by unavoidable delay. Interestingly, MSM methods also contain a small but temperature-dependent spurious threshold-voltage shift caused by mobility degradation but appear to be undistorted by the interface state dynamics. OTF methods, on the other hand, suffer from the problem of the initial measurement and can be polluted by rather complex mobility changes. Furthermore, all OTF methods rely on simple compact models for the back-extrapolation of the threshold-voltage shift at stress-level, which results in difficult-to-quantify inaccuracies.

Overall, it could be confirmed that all measurement techniques used to determine the degradation after bias temperature stress are prone to errors. Nevertheless, in spite of the inherent delay, the smaller systematic errors in the conventional MSM technique allow one to accurately monitor the time evolution of the recovery, which can then be used for the formulation of physics-based models.

Acknowledgment

The authors would like to thank H. Reisinger, V. Huard, S. Mahapatra, and N. Mielke for the fruitful discussions.

REFERENCES

[1] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor and circuit: Models, mechanisms and scaling effects [MOSFETs]," in *IEDM Tech. Dig.*, 2003, pp. 14.5.1–14.5.4.

[2] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, Jan. 2006.

[3] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy?" in *Proc. IRPS*, 2007, pp. 1–9.

[4] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *Proc. IRPS*, 2005, pp. 381–387.

[5] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, "Analysis of NBTI degradation- and recovery-behavior based on ultra fast $V_{\rm th}$-measurements," in *Proc. IRPS*, 2006, pp. 448–453.

[6] C. Shen, M.-F. Li, X. P. Wang, Y.-C. Yeo, and D.-L. Kwong, "A fast measurement technique of MOSFET $I_{\rm D} - V_{\rm G}$ characteristics," *IEEE Electron Device Lett.*, vol. 27, no. 1, pp. 55–57, Jan. 2006.

[7] D. Heh, R. Choi, C. D. Young, B. H. Lee, and G. Bersuker, "A novel bias temperature instability characterization methodology for high-k nMOSFETs," *IEEE Electron Device Lett.*, vol. 27, no. 10, pp. 849–851, Oct. 2006.

[8] T. Grasser and B. Kaczer, "Negative bias temperature instability: Recoverable versus permanent degradation," in *Proc. ESSDERC*, 2007, pp. 127–130.

[9] T. Grasser, B. Kaczer, P. Hehenberger, W. Gos, R. O'Connor, H. Reisinger, W. Gustin, and C. Schlünder, "Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability," in *IEDM Tech. Dig.*, 2007, pp. 801–804.

[10] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac, and N. Revil, "On-the-fly characterization of NBTI in ultra-thin gate oxide pMOSFETs," in *IEDM Tech. Dig.*, 2004, pp. 109–112.

[11] J. F. Zhang and M. H. Chang, "An assessment of effective mobility variation during negative bias temperature instability," *ECS Trans.*, vol. 6, no. 3, pp. 301–311, 2007.

[12] H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin, and C. Schlünder, "A comparison of fast methods for measuring NBTI degradation," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 4, pp. 531–539, Dec. 2007.

[13] A. Neugroschel, G. Bersuker, R. Choi, C. Cochrane, P. Lenahan, D. Heh, C. Young, C. Y. Kang, B. H. Lee, and R. Jammy, "An accurate lifetime analysis methodology incorporating governing NBTI mechanisms in high-k/SiO_2 gate stacks," in *IEDM Tech. Dig*, 2006, pp. 1–4.

[14] M. A. Alam, "A review of characterization methodologies of gate dielectric breakdown and negative bias temperature instability," in *Proc. IPFA*, Jul. 2006, pp. 25–32.

[15] Y. Tsividis, *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1999.

[16] S. E. Rauch, III, "The statistics of NBTI-induced $V_{\rm T}$ and β mismatch shifts in pMOSFETs," *IEEE Trans. Device Mater. Rel.*, vol. 2, no. 4, pp. 89–93, Dec. 2002.

[17] L.-A. Ragnarsson and P. Lundgren, "Electrical characterization of P_b centers in (100)Si-SiO_2 structures: The influence of surface potential on passivation during post metallization anneal," *J. Appl. Phys.*, vol. 88, no. 2, pp. 938–942, Jul. 2000.

[18] T. Grasser, R. Entner, O. Triebl, H. Enichlmair, and R. Minixhofer, "TCAD modeling of negative bias temperature instability," in *Proc. SISPAD*, Monterey, CA, Sep. 2006, pp. 330–333.

[19] J. Bastos, M. Steyaert, A. Pergoot, and W. Sansen, "Mismatch characterization of submicron MOS transistors," *Analog Integr. Circuits Signal Process.*, vol. 12, no. 2, pp. 95–106, Feb. 1997.

[20] T. Grasser, W. Goes, V. Sverdlov, and B. Kaczer, "The universality of NBTI relaxation and its implications for modeling and characterization," in *Proc. IRPS*, 2007, pp. 268–280.

[21] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, C. Guerin, G. Ribes, F. Perrier, M. Mairy, and D. Roy, "Paradigm shift for NBTI characterization in ultra-scaled CMOS technologies," in *Proc. IRPS*, 2006, pp. 735–736.

[22] *MINIMOS-NT 2.1 User's Guide*, Institut für Mikroelektronik, Technische Universität Wien, Wien, Austria, 2004. IμE. [Online]. Available: www.iue.tuwien.ac.at

[23] C. Jungemann, T. Grasser, B. Neinhüs, and B. Meinerzhagen, "Failure of moments-based transport models in nanoscale devices near equilibrium," *IEEE Trans. Electron Devices*, vol. 52, no. 11, pp. 2404–2408, Nov. 2005.

[24] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Density of states and structure of NBTI-induced defects in plasma-nitrided pMOSFETs," in *Proc. IRPS*, 2007, pp. 503–510.

[25] V. M. Agostinelli, Jr., H. Shin, and A. F. Tasch, Jr., "A comprehensive model for inversion layer hole mobility for simulation of submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 38, no. 1, pp. 151–159, Jan. 1991.

[26] A. Haggag, G. Anderson, S. Parihar, D. Burnett, G. Abeln, J. Higman, and M. Moosa, "Understanding SRAM high-temperature-operating-life NBTI: Statistics and permanent vs recoverable damage," in *Proc. IRPS*, 2007, pp. 452–456.

[27] A. E. Islam, E. N. Kumar, H. Das, S. Purawat, V. Maheta, H. Aono, E. Murakami, S. Mahapatra, and M. A. Alam, "Theory and practice of on-the-fly and ultra-fast VT measurements for NBTI degradation: Challenges and opportunities," in *IEDM Tech. Dig.*, 2007, pp. 805–808.

[28] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, "Recent issues in negative-bias temperature instability: Initial degradation, field dependence of interface trap generation, hole trapping effects, and relaxation," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2143–2154, Sep. 2007.

[29] D. Bauza, "Rigorous analysis of two-level charge pumping: Application to the extraction of interface trap concentration versus energy profiles in metal–oxide–semiconductor transistors," *J. Appl. Phys.*, vol. 94, no. 5, pp. 3239–3248, Sep. 2003.

[30] D. Goguenheim, D. Vuillaume, G. Vincent, and N. M. Johnson, "Accurate measurements of capture cross sections of semiconductor insulator interface states by a trap-filling experiment: The charge-potential feedback effect," *J. Appl. Phys.*, vol. 68, no. 3, pp. 1104–1113, Aug. 1990.

[31] T. L. Tewksbury and H.-S. Lee, "Characterization, modeling, and minimization of transient threshold voltage shifts in MOSFETs," *IEEE J. Solid-State Circuits*, vol. 29, no. 3, pp. 239–252, Mar. 1994.

[32] C. Shen, M.-F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast $V_{\rm th}$ transient in NBTI of pMOSFETs with SiON dielectric," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

Tibor Grasser (M'05–SM'05) was born in Vienna, Austria, in 1970. He received the Diplomingenieur degree in communications engineering, the Ph.D. degree in technical sciences, and the Venia Docendi in microelectronics from the Technische Universität (TU) Wien, Wien, Austria, in 1995, 1999, and 2002, respectively.

He is currently an Associate Professor at the Institute for Microelectronics, TU Wien. Since 1997, he has headed the MINIMOS-NT development group, working on the successor of the highly successful MiniMOS program. He was a Visiting Research Engineer at Hitachi Ltd., Tokyo, Japan, and at the Alpha Development Group, Compaq Computer Corporation, Shrewsbury, MA. Since 2003, he has been the Head of the Christian Doppler Laboratory for TCAD in Microelectronics, an industry-funded research group embedded in the Institute for Microelectronics, TU Wien. He is the coauthor or author of over 200 articles in scientific books, journals, and conference proceedings and is the Editor of a book on advanced device simulation. His current scientific interests include circuit and device simulation, device modeling, and reliability issues.

Dr. Grasser has been involved in the program committees of conferences such as SISPAD, IWCE, ESSDERC, IRPS, IIRW, and ISDRS. He was also a Chairman of SISPAD 2007.

Paul-Jürgen Wagner was born in Vienna, Austria, in 1979. He received the Diplomingenieur degree in electrical engineering from the Technische Universität Wien, Wien, Austria, in 2007, where he is currently working toward the Ph.D. degree in the Christian Doppler Laboratory for TCAD in Microelectronics, Institute for Microelectronics.

Philipp Hehenberger was born in Vienna, Austria, in 1980. He received the Diplomingenieur degree in technical physics from the Technische Universität (TU) Wien, Wien, Austria, in 2006. He is currently working toward the Ph.D. degree with the Institute for Microelectronics, TU Wien.

In 2008, he was a visitor with the School of Microelectronics, Fudan University, Shanghai, China. His current scientific interests include device modeling of HCI, NBTI, and advanced charge-pumping techniques.

Wolfgang Goes was born in Vienna, Austria, in 1979. He received the Diplomingenieur degree in technical physics from the Technische Universität (TU) Wien, Wien, Austria, in 2005, where he is currently working toward the Ph.D. degree in the Christian Doppler Laboratory for TCAD in Microelectronics, Institute for Microelectronics, since 2006. His diploma thesis addressed grain boundaries in back contact solar cells.

In 2007, he was a Visitor at the Department of Physics and Astronomy, Vanderbilt University, Nashville, TN. His current scientific interests include first-principle simulations of the chemical processes involved in NBTI and HCI, modeling of tunneling and charge trapping in dielectrics, and reliability issues in general.

Ben Kaczer received the M.S. degree in physical electronics from Charles University, Prague, Czech Republic, in 1992 and the M.S. and Ph.D. degrees in physics from the Ohio State University, Columbus, in 1996 and 1998, respectively.

Since 1998, he has been with the Reliability Group, IMEC, Leuven, Belgium, where his activities have included the research of the degradation phenomena and reliability assessment of SiO2, SiON, high-k, and ferroelectric films, planar and multiple-gate FETs, circuits, and characterization of Ge and III-V devices. He has authored or coauthored more than 100 journal and conference papers and presented eight invited presentations and two IRPS tutorials.

Dr. Kaczer has served or is serving in various functions at the IEDM, IRPS, SISC, and INFOS conferences. He is the recipient of the Best and the Outstanding Paper Awards at IRPS. For his Ph.D. research on the ballistic-electron emission microscopy of SiO2 and SiC films, he received the OSU Presidential Fellowship and support from Texas Instruments, Inc.