Modeling of Hot-Carrier Degradation: Physics and Controversial Issues

Stanislav Tyaginov and Tibor Grasser
*Institute for Microelectronics, Vienna University of Technology, Gußhausstraße 27-29, A-1040 Wien, Austria
Email: tyaginov|grasser@iue.tuwien.ac.at

***Abstract*—We discuss and analyze the main features of hot-carrier degradation (HCD) which are a strong localization at the drain-side of the device, the interplay between single- and multiple-particle processes of Si-H bond dissociation, the transition of the worst-case scenario when going from long- to short-channel devices, and its temperature dependence. These main peculiarities are then linked to the physical processes responsible for HCD. We show that the problem can be conditionally separated into three main subtasks: the carrier transport aspect, the kinetics of defect generation, and modeling of the degraded devices. From this perspective, the most important physics-based models and their validity are discussed. In order to obtain a most accurate description of HCD, we try to minimize the number of empirical parameters by basing our own model on a thorough treatment of carrier transport. Finally, we discuss one of the most important open obstacles towards the understanding of HCD, namely whether bulk oxide traps contribute to the damage or not.**

I. Introduction

If a source-drain voltage is applied to a MOSFET, the carriers in the channel are accelerated by the electric field and can gain a substantial amount of energy, making them "hot carriers", see Fig. 1. As these hot carriers impinge onto the interface, they can produce damage which results in a degradation of the transistor performance. This degradation mode has been known for almost five decades [1] and is called hot-carrier degradation (HCD). The damage is typically explained by the generation of interface traps [2], [3]. In addition, bulk oxide traps and processes related to capture and emission of carriers by these traps also contribute to the degradation [3]–[5].

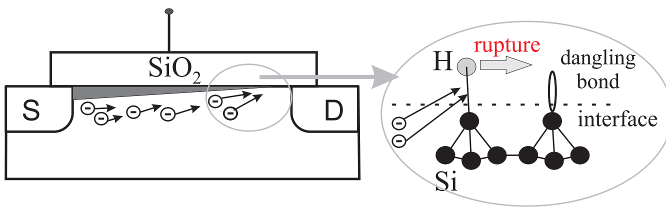

Fig. 1. A schematic presentation of hot-carrier degradation. The dissociation of the Si-H bond induced by the successive bombardment of two hot carriers is sketched in the right part.

Interface traps are created by dissociation of the Si-H bonds at the $\mathrm{Si/SiO_2}$ interface. This interface is a non-regular system and characterized by disorder [6]. Among other issues, this results in dangling bonds at the interface. These dangling bonds can capture carriers, become charged, and thus degrade the device performance. To avoid this, hydrogen is intentionally incorporated into the system to passivate these bonds.

Unfortunately, these passivated bonds can be broken as a consequence of hot carrier bombardment. More energetical carriers have a higher probability to trigger the rupture event. As the carrier energy increases on average towards the drain end of the device, the density of the interface traps (N_{it}) is a distributed quantity and depends on the coordinate x along the $\mathrm{Si/SiO_2}$ interface, see Fig. 2. The density N_{it} also varies with energy across the silicon bandgap. The interface traps generated during the stress can become charged by capturing carriers. As a result, the distortion of the device electrostatics is induced. This distortion is observable as a shift of the threshold voltage. Moreover, the charged defects distributed over the device degrade the mobility and the device characteristics such as the transconductance G_{m} and the linear drain current I_{dlin}.

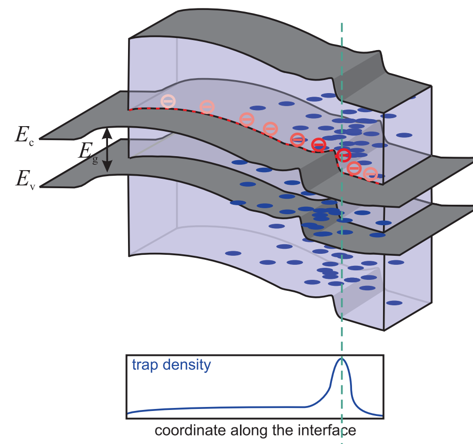

Fig. 2. A MOSFET under hot-carrier stress and the schematic visualization of the density $N_{\mathrm{it}}(x)$. Charged traps induce a distortion of the device electrostatics as well as a mobility degradation. The density N_{it} varies with the coordinate and in energy.

Since the driving force of the carrier acceleration in the channel is the electric field, it is often assumed that the maximum of the interface state generation rate just corresponds to the electric field peak [7]. One of the first successful HCD models was the so-called "lucky-electron" model proposed by Hu [7]. A little later, the IBM group demonstrated that the interface state generation probability depends only on the carrier energy rather than on the oxide field and is insensitive to the stress mechanism [8], [9]. This suggests HCD to be an energy driven phenomenon.

This controversy can be reconciled considering Si-H bond-breakage as a stochastic process which can be described in terms of a probability that a carrier characterized by a certain energy can break a bond. Therefore, for a proper description of the bond dissociation one should consider not a solitary carrier interacting with a bond but the whole carrier ensemble. The particles in this ensemble can have substantially different energies and each carrier makes a contribution to the

 978-1-4673-2752-7/12/$31.00 ©2012 IEEE

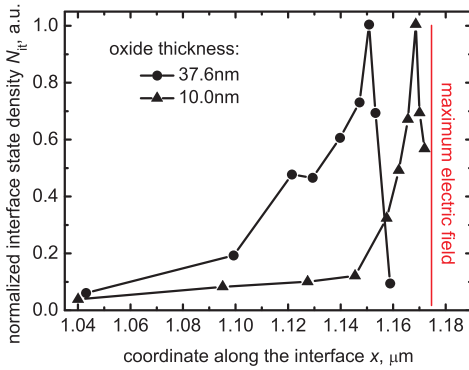

Fig. 3. The interface state density as a function of the lateral coordinate along the interface extracted from charge-pumping data for two MOSFETs with different oxide thicknesses. Both $N_{\rm it}(x)$ profiles demonstrate rather sharp maxima near the electric field peak. The data are from [2].

entire bond dissociation process. Therefore, a proper physics-based description of HCD requires self-consistent treatment of carrier transport and bond dissociation kinetics. Eventually, however, the most prominent manifestation of HCD is a change of the device characteristic during stress. Therefore, for proper understanding and modeling of hot-carrier damage it is essential to link the microscopic mechanisms of defect generation with the device characteristics. These various aspects of the phenomenon make the problem very complex.

In the next section the main peculiarities of HCD will be described and linked to the physical mechanisms discussed in Section 3. Then we present the most important physics-based concepts of HCD modeling (including our own approach) which attempt to properly cover these features and mechanisms. In Section 5 some open issues in understanding and modeling of this detrimental phenomenon will be analyzed.

II. Main Peculiarities of HCD

One of the main peculiarities of HCD is its strong localization near the pinch-off region at the drain end of the gate, which is close to the area where the electric field peaks. This is because carriers need to gain a substantial amount of energy in order to trigger Si-H bond dissociation. One of the pioneering works devoted to the lateral distribution of the interface state density is due to Ancona *et al.* [2]. They used the charge pumping technique to resolve $N_{\rm it}$ as a function of the coordinate along the interface. The $N_{\rm it}$ profiles for transistors with two different oxide thicknesses extracted from charge-pumping data are plotted in Fig. 3 [2]. In both cases $N_{\rm it}(x)$ demonstrates a rather sharp peak located near the electric field maximum.

This fact is reflected in many modeling attemps. For instance, the "lucky-electron" model states that the driving force of HCD is the electric field. However, in the eighties device dimension shrinking was faster than the power supply reduction which was leading to higher electric fields [10], [11]. As a result, hot-carrier degradation was severe in those devices demanding a reduction of the supply voltage faster than the device size. In modern ultra-scaled MOSFETs, the supply voltage is therefore scaled down to 1.0-1.5V while

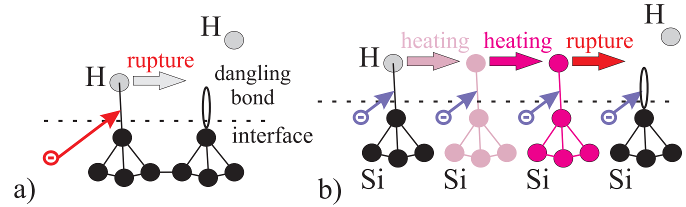

Fig. 4. A schematic representation of the SP- and MP-mechanisms. According to the SP-process a solitary energetical carrier can dissociate the bond. The MP-mechanism corresponds to the subsequent bombardment of the the bond by several colder carriers followed by the bond excitation and eventually the H release.

the threshold voltage of the Si-H bond dissociation process is ~3eV. As a consequence, it was expected that HCD would be totally suppressed in ultra-scaled devices [10], [12]. However, in practice even transistors with channel lengths of less than 150nm and/or subjected to hot-carrier stress with the source-drain voltage $V_{\rm ds}$ below 1.5V can suffer severely from HCD [13]–[15].

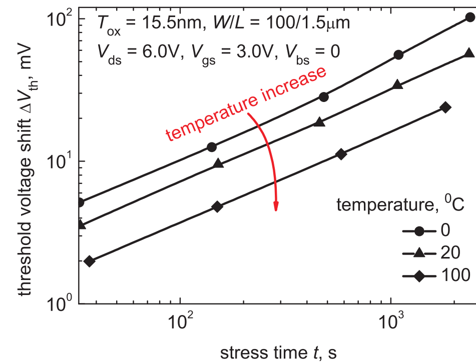

Fig. 5. The threshold voltage shift measured vs. the stress time at three different temperatures in the device stressed at the worst-case conditions (data from [16]). One can see that $\Delta V_{\rm th}$ becomes less pronounced as temperature increases.

One of the reasons why hot-carrier degradation can be severe in ultra-scaled devices is that carriers can gain energy through exchange mechanisms. Another factor is that the energy for Si-H bond dissociation can be provided by subsequent collisions with multiple carriers. In a long-channel and/or high-voltage device, carriers bombarding the interface are already rather energetical, thereby triggering Si-H bond rupture by a single collision, which is referred to as the single-particle (SP) mechanism. Such extremely hot carriers are unlikely in scaled devices. Instead, several colder particles subsequently bombard a bond, thereby exciting and eventually rupturing it, which is referred to as the multiple-particle (MP) process. As a consequence, the dominant mechanism of bond dissociation changes when device dimensions shrink [12], [17]–[20].

The interplay between the SP- and MP-mechanisms explains why the worst-case conditions of HCD are different in long- and short-channel devices. In particular, in a long-channel/high-voltage n-MOSFET the worst-case scenario of HCD corresponds to the peak substrate current $I_{\rm sub}$, or – in other words – to the maximum impact ionization rate [21], [22]. The $I_{\rm sub}$ maximum is typically observed when

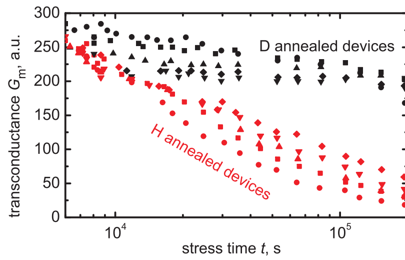

Fig. 6. The transconductance change during to hot-carrier stress in hydrogen and deuterium annealed devices. In latter ones the degradation is lower for all stress times, data from [27].

$V_{\mathrm{gs}} \approx 0.5V_{\mathrm{ds}}$. As for long-channel p-MOSFETs, the worst HCD corresponds to the maximum gate current I_{g} and such a simple empirical interrelation between the voltages cannot be established [23], [24]. In general, the most severe HCD in long-channel devices is realized at the maximum average carrier energy, i.e. when the carrier ensemble contains a substantial fraction of hot carriers.

In scaled devices, however, these carrier do not exist in sufficient numbers. As an alternative, the MP-mechanism becomes important which requires only low carrier energies. However, the bombardment of the interface by these colder carriers needs to be intensive. As a result, the carrier flux rather than the carrier energy becomes important. Since the maximum carrier flux in ultra-scaled devices occurs when $V_{\mathrm{gs}} = V_{\mathrm{ds}}$ in both n- and p-channel devices, this regime corresponds to the worst-case scenario of HCD [20], [25], [26].

Another important attribute of HCD is its temperature dependence. One of the pioneering papers devoted to this issue is due to Hsu and Chiu [16]. A long-channel n-MOSFET was stressed at V_{gs}=3.0 and V_{ds}=6.0V (worst-case conditions) and three different temperatures of T = 0, 20, and 100°C. Fig. 5 demonstrates that ΔV_{th} degrades less at higher temperatures for all stress times. This is the typical behavior in long-channel devices. In ultra-scaled devices, on the other hand, HCD becomes more severe at elevated temperatures [28]–[30]. This is because hot electrons are created by electron-electron scattering, which plays a crucial role in ultra-scaled devices [31]–[34].

Finally, the group of Hess demonstrated that if instead of hydrogen deuterium is used to passivate the dangling bonds at the $\mathrm{Si/SiO_2}$ interface, the MOSFET is much more resistive with respect to hot-carrier stress [27]. Fig. 6 demonstrates that the mobility degrades much less in deuteriorated devices as compared to their hydrogenated counterparts. It was thus concluded that the Si-D bond is more robust than the Si-H bond with respect to the hot-carrier bombardment.

III. Physical Mechanisms responsible for HCD

The peculiarities of hot-carrier degradation suggest that HCD is defined by the way carriers in the ensemble are distributed over energy and by the strength of the Si-H/Si-D bond. A carrier interacting with the interface can induce a bond dissociation and the kinetics of this process depend on the carrier energy and on the activation energy for bond dissociation. The probability of finding a particle in the energy range $[E; E+\Delta E]$ is given by the carrier energy distribution function (DF). The interplay between the SP- and MP-mechanisms stems from the fractions "hot" and "colder" carriers. At the same time, SP- and MP-processes are linked to different quantum processes which are defined by the bond binding energy and the spectrum of vibrational modes. The localization of HCD is explained as a trade-off between the carrier flux and carrier energy and is again defined by carrier transport combined with the bond energetics. The temperature behavior of HCD is defined by scattering mechanisms and for their proper description the Boltzmann transport equation has to be solved. Therefore, carrier transport and the bond energetics are two important aspects of the problem.

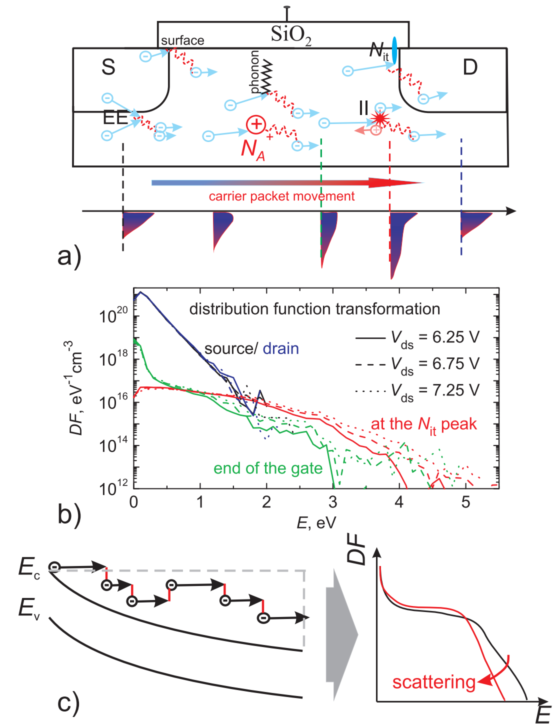

Fig. 7. A schematic representation of the scattering mechanisms in MOSFETs (a), the evolution of the carrier distribution function with the coordinate along the interface (b), and the suppression of the hot-carrier fraction of the statistical ensemble (c).

A. The Carrier Transport Aspect

While traversing the channel, carriers gain energy from the electric field, see Fig. 7. On the other hand, they also experience scattering events and exchange energy with other particles, such as phonons and other carriers. The most important scattering mechanisms affecting carrier transport are electron-phonon scattering, scattering at ionized impurities, impact ionization, surface scattering, and electron-electron scattering, see Fig. 7a. In degraded devices, where interface states have already been generated, trapped charges lead to additional scattering. Due to these scattering mechanisms, the high energetical fraction of the statistical ensemble is depopulated and the high-energy tail of the distribution function

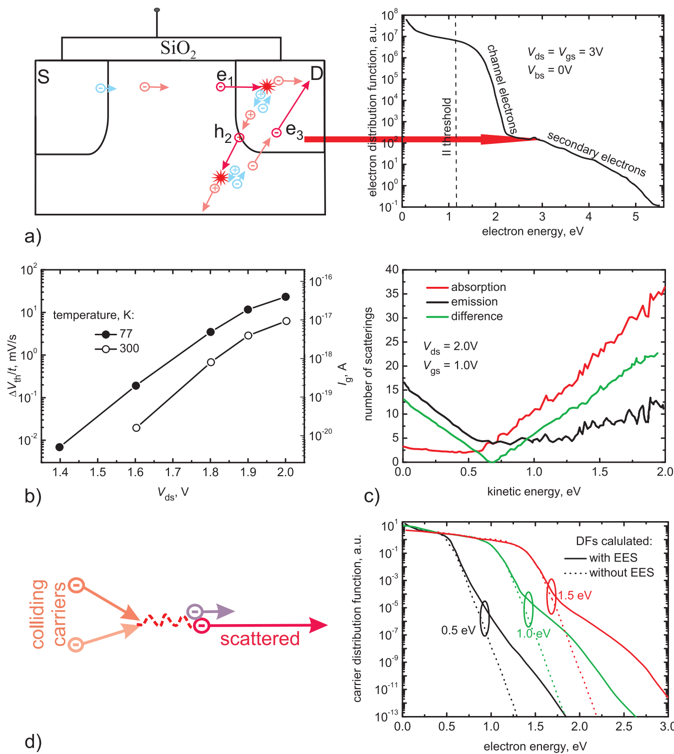

Fig. 8. Mechanisms populating the high-energy tails of the carrier DF in ultra-scaled devices. The impact ionization feedback via the high vertical electric field at the drain-substrate junction (a) [15]. The Auger recombination which leads to tunneling of hot carriers and the threshold voltage shift (b) [13]. Electron-phonon scattering generates hot carriers if the number of absorbed phonons exceeds the number of emitted ones (c) [35]. Electron-electron scattering results in DFs propagating far beyond the energy available from the electric field (d) [31].

is distorted, see Fig. 7c. This tendency also explains the suppression of HCD at elevated temperatures. In fact, the rate of all the scattering mechanisms increases with temperature. Therefore, the suppression of the high-energy tail of the DF, and hence suppression of HCD becomes more efficient. Quite to the contrary, in ultra-scaled devices electron-electron scattering is the dominant mechanism, which populates the hot carrier fraction more efficiently at elevated temperatures, making HCD more severe.

The evolution of the DF along the interface is shown in Fig. 7b, see [36]. As an example we used a 5V n-MOSFET with a channel length of 0.5μm. One can see that near the source carriers are in equilibrium and follow a Maxwellian distribution. In contrast, in the middle of the channel and especially closer to the drain end of the gate (where the electric field peaks) the carrier distribution is severely non-Maxwellian. For instance, the red curves which correspond to the position near the $N_{\rm it}$ peak show quite long high-energy tails and a plateau at moderate energies. If we shift further towards the drain, a Maxwellian part is pronounced while high-energy tails become shorter. This transformation is related to the influence of the drain which is a reservoir of thermalized carriers. One can also see that at higher $V_{\rm ds}$ high-energy tails of the DF are more populated.

However, scattering mechanisms can populate the high-energy fraction of the carrier ensemble in ultra-scaled devices, see Fig. 8 [13], [15], [31], [35]. In [15] an n-MOSFET fabricated on a 0.1 CMOS design was investigated using a Monte-Carlo device simulator. $I_{\rm g}$ was used as a sensitive measure of the high-energy tail of the carrier distribution function because hot electrons can tunnel more efficiently than thermalized ones. The mechanism responsible for $I_{\rm g}$ is based on the impact ionization feedback through the vertical fields of the drain-substrate junction (right panel of Fig. 8a). Impact ionization generates electron-hole pairs and the secondary carriers are then accelerated by the electric field, thereby contributing to the gate current and high-energy tails of the DF.

Another mechanism which forms high-energy tails in the DF is Auger recombination. n-MOSFETs with a channel length of 0.3μm were subjected to hot-carrier stress at $V_{\rm ds} \geq$1.4V [13]. The oxide thickness was 45nm and the gate current was monitored to assess how hot the carriers are since tunneling of thermalized carriers is totally suppressed in this thick film, see Fig. 8b. At $V_{\rm ds} \geq 1.4$V impact ionization already comes into play. The concentrations of electrons and holes near the drain are substantial and carrier recombination cannot be neglected. Hot carriers are thereby generated by Auger recombination due to transferring of the energy of two recombining carriers to another electron which becomes hot.

In addition, an electron can gain energy from phonons if the number of absorbed phonons exceeds the number of emitted ones, Fig. 8c. Such a scenario was supported by simulations using a full-band Monte-Carlo device simulator in [35]. The authors have demonstrated that this process populates the high-energy tail of the carrier DF.

Finally, electron-electron scattering is the dominating mechanism populating the high energetical fraction of the carrier ensemble in ultra-scaled devices, see Fig. 8d. In [31] the authors solved a one-dimensional Boltzmann transport equation extended in order to include electron-electron interactions. The right plot of Fig. 8d shows the DFs calculated with and without electron-electron scattering for different energies available from the electric field: 0.5, 1.0, and 1.5eV. One can see that electron-electron scattering dramatically changes the shape of the DF which propagates to energies higher that those available form the electric field.

B. *Energetics of the Si-H Bond*

Si-H bond dissociation leads to the creation of electrically active defects which degrade the performance of CMOS devices and contribute not only to HCD but also to the relative phenomenon, the bias temperature instability (BTI) [37]–[39]. As a result, the energetics of the Si-H bond has been the subject of intensive research. Fig. 9 obtained using *ab initio* calculations with density functional theory demonstrates the pathway of the hydrogen release reaction. The pathway connects the stable configuration corresponding to the Si-H bond and the transport mode. One can see that the potential barrier separating these states is ~2.0eV, which explains why we are primarily interested in *hot carriers*: for the bond

dissociation one should deliver to the H atom a bond-breaking portion of energy of above 2eV. Just particles with energies above 2eV can effectively trigger the bond dissociation event through the SP-mechanism.

The disparity between the electron mass and the mass of the hydrogen nucleus is about 1000. The total momentum of the system has to be conserved. Therefore, a direct bombardment of the hydrogen nucleus by an electron will lead to a negligible portion of energy transferred to H, and the bond dissociation is unlikely. Hence, it has been suggested that the most probable way to dissociate the bond is via excitation of one of the bonding electrons to an anti-bonding state [12]. As a consequence, a repulsive force acting on the H atom is induced, followed by the release of hydrogen.

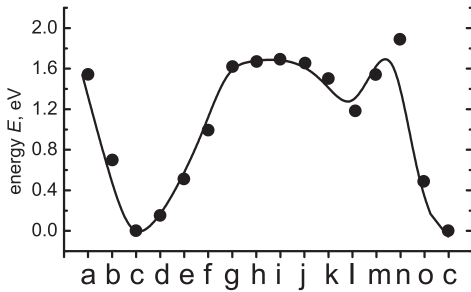

Fig. 9. The pathway of the H release reaction which connects the bonded state (corresponding to the Si-H bond) and the transport mode. The potential barrier separating these states is above 2eV. Data are from [40].

An alternative mechanism of bond dissociation is the MP-process which is responsible for HCD in scaled devices. The physical mechanism behind this process is multiple vibrational bond excitation due to the bombardment of the bond by several *colder carriers* [41]–[44]. The first success of the multivibrational concept was achieved when hydrogen/deuterium desorption induced by subsequent bombardment by several (cold) carriers tunneling from an STM tip was investigated [41], [42]. These experiments showed that the D-passivated surfaces are much more resistant with respect to electron bombardment as compared to hydrogenated ones. The difference in depassivation rates can be more than two orders of magnitude at high voltages [43], which gave rise to the name "giant isotope effect".

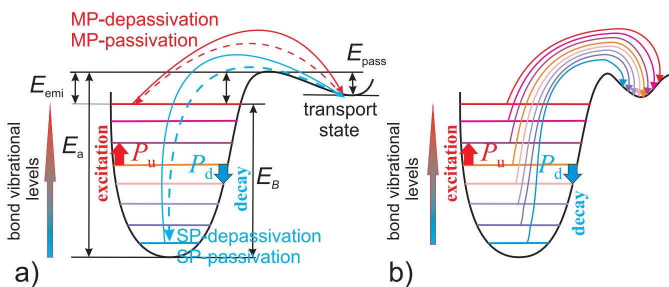

Fig. 10. The Si-H bond as a truncated harmonic oscillator. The SP- and MP-mechanisms correspond to the H release from the ground and the last bonded level, respectively (a). In the Hess model [19] hydrogen hopping from all the intermediate states was considered (b).

To model the dissociation, the Si-H bond is treated as a truncated harmonic oscillator characterized by the system of eigenstates, see Fig. 10. The hydrogen excitation from the last bonded level to the transport state corresponds to bond rupture. The electron flux can induce either phonon absorption (that is, bond heating) or phonon emission (related to the multivibrational mode decay).

Simulations trying to explain the giant isotope effect have been performed in [45]. The authors simulated vibrational excitation using a tight-binding molecular dynamics method. Both Si-H and Si-D bonds have stretch and bending vibrational modes. The stretch mode of Si-H and Si-D bonds is very stable demonstrating no decay within 1ns. As for the bending mode, the situation is dramatically different: in the case of the Si-H bond the excitation does not decay within first 0.8ps, while the Si-D bending mode decreases rapidly within 1-2 oscillations. Since at Si/SiO_2 interfaces the H/D release occurs via bond-bending distortions, these calculations could explain the D stability at the interfaces.

IV. Physics-based models of HCD

A. Hess Model

The revolutionary contribution of Hess *et al.* to hot-carrier degradation modeling was to link carrier transport and the energetics of the Si-H bond [12], [19], [46]. This made it possible to understand the two competing mechanisms of bond dissociation: SP- and MP-processes. The interplay between these processes is controlled by fractions of "hot" and "colder" carrier. Thus, Hess *et al.* first expressed the idea that HCD is controlled by the distribution function or a quantity based on the DF, the *carrier acceleration integral* (AI).

The acceleration integral represents the desorption rate via excitation of one of the bonding electrons to an anti-bonding state (the SP-process). The AI is defined as

$$R_{\mathrm{SP}} \sim \int_{E_{\mathrm{th}}}^{\infty} F(E)P(E)\sigma(E)\mathrm{d}E, \qquad (1)$$

where $F(E)$ is the carrier impact frequency on the surface per unit area and energy (i.e. the product of the DF, density-of-states, and the carrier velocity), $\sigma(E)$ is the scattering cross section for excitation of the bonded electron, $P(E)$ is the probability that such an excitation leads to H desorption, while E_{th} is the threshold energy for scattering.

As for the MP-process, the Si-H bond is represented by a truncated harmonic oscillator, Fig. 10a. The multivibrational modes are characterized by the corresponding phonon lifetime and the distance between the levels. The occupation number obeys a Bose-Einstein distribution. The rates P_u and P_d of bond heating/cooling induced by the electron flux are described by the same constructions as the acceleration integral [46]:

$$P_{\mathrm{d}} \sim \int_{E_{\mathrm{th}}}^{\infty} I(E)\sigma_{\mathrm{ab}}(E)[1-f_{\mathrm{ph}}(E-\hbar\omega)]\mathrm{d}E,$$
$$P_{\mathrm{u}} \sim \int_{E_{\mathrm{th}}}^{\infty} I(E)\sigma_{\mathrm{emi}}(E)[1-f_{\mathrm{ph}}(E+\hbar\omega)]\mathrm{d}E, \qquad (2)$$

where $\sigma_{\mathrm{ab}}(E)/\sigma_{\mathrm{emi}}(E)$ are the phonon absorption/emission reaction cross sections, $\hbar\omega$ the phonon energy, and $f_{\mathrm{ph}}(E)$ the phonon occupation numbers. The bond-breakage rate R_{MP} of the MP-process is:

$$\left(\frac{E_{\mathrm{B}}}{\hbar\omega}+1\right)\left[P_{\mathrm{d}}+\exp\left(-\frac{\hbar\omega}{k_{\mathrm{B}}T_{\mathrm{L}}}\right)\right] \times \left[\frac{P_{\mathrm{u}}+\omega_{\mathrm{e}}}{P_{\mathrm{d}}+\exp(-\hbar\omega/k_{\mathrm{B}}T_{\mathrm{L}})}\right]^{-E_{\mathrm{B}}/\hbar\omega}, \qquad (3)$$

with E_{B} being the energy of the last bonded level in the quantum well (Fig. 10a) and the phonon reciprocal life-time ω_{e}.

A further refinement of the Hess model was obtained by incorporating the H hopping from all the levels in the quantum well [19]. Following the work of Perrson and Avouris [42], the authors simplified the rates (2) by linking them with the drain current I_{d}. Finally, the bond-breakage rate is obtained as

$$R=\sum_{i=0}^{N_{\mathrm{l}}}\left[\frac{I_{\mathrm{d}}f_{\mathrm{v}}+\omega_{\mathrm{e}}\exp\left(-\hbar\omega/k_{\mathrm{B}}T_{\mathrm{L}}\right)}{I_{\mathrm{d}}f_{\mathrm{v}}+\omega_{\mathrm{e}}}\right]A^{\mathrm{i}}I_{\mathrm{d}}f_{\mathrm{d}}, \qquad (4)$$

where the first part of the ith term defines the population of the ith level (dependent on the AI). The second part describes the hopping from this level.

Another characteristic feature of the Hess model is the assumption that the activation energy of bond-breakage is statistically distributed [17]. This assumption is supported by *ab initio* calculations [40]. The dispersion of the activation energy leads to different power-law slopes during degradation, see [17]. This idea is also confirmed by experimental observations.

Despite the pioneering nature of the Hess model, there are a few of shortcomings resulting from the way the model was simplified to bring it into a compact form. First of all, although the interface traps are considered at the microscopic level, they remain unconnected to the device level. Thus, the device life-time is estimated as the time when the concentration N_{it} reaches a certain level and the degradation of such *device* parameters as transconductance, linear drain current is not really addressed. Finally, although the necessity of evaluation of the carrier DF is acknowledged, in practice this information has not been incorporated into the approach.

B. Energy Driven Paradigm by Rauch and LaRosa

An alternative empirical model was suggested by Rauch and LaRosa [32], [34], [47], who acknowledged the importance of the carrier energy on the degradation, and thus this model is called "energy-driven paradigm". The energy-driven paradigm claims that beyond the 180 nm node the fundamental driving

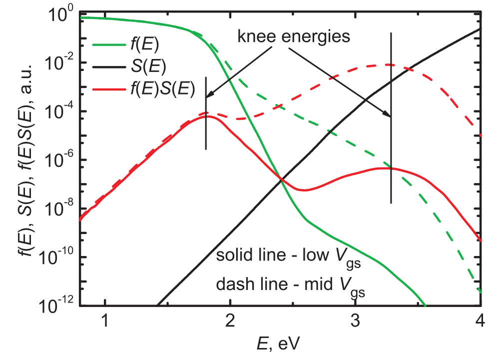

Fig. 11. The carrier DF $f(E)$, the reaction cross section $S(E)$, and their product. The trade-off between rapidly decaying $f(E)$ and the polynomial increase of $S(E)$ results in a sharp maximum of the integrand in the acceleration integral pronounced at the "knee energy".

force of HCD changes from the maximum electric field ("lucky-electron" model) to the energy deposited by carriers. The two hot-carrier driven phenomena impact ionization and interface state generation are described by integrals of the form $\int f(E)S(E)\mathrm{d}E$. Here $f(E)$ is the DF and $S(E)$ the reaction cross section. The DF is a strongly decaying function of energy, while $S(E)$ has a power-law energy dependence. This trade-off results in a sharp maximum of the integrand pronounced at a certain energy, see Fig. 11. This energy is called "knee energy" and is a weak function of the applied bias V_{ds}.

The main advantage of this model is that time-consuming calculations of the carrier DF are avoided and the information about the distribution is represented using an empirical parameter. Although this paradigm substantially simplifies the treatment of HCD, it suffers from some shortcomings. First, the model does not consider the lateral profile $N_{\mathrm{it}}(x)$ and the localized nature of the damage is thus not represented. Following the Hess approach, the device life-time is estimated by the interface state generation rate. However, it would be more reasonable to define it as the time when the change of some device characteristic reaches a critical value.

C. Bravaix Model

The model of Bravaix *et al.* [20], [48], [49] inherits the main features of both the Hess and the Rauch/LaRosa approaches: the interplay between SP- and MP-mechanisms and the idea that HCD is defined by the carrier DF. Similarly to the Rauch/LaRosa paradigm, in the Bravaix model calculations of the DF are substituted by operation/stress condition related empirical factors. For instance, the rates $P_{\mathrm{u}}/P_{\mathrm{d}}$ for bond heating/relaxation are linked to the drain current via the empirical factor S_{SP}:

$$P_{\mathrm{d}}=S_{\mathrm{MP}}(I_{\mathrm{d}}/e)+\omega_{\mathrm{e}}\exp\left(-\hbar\omega/k_{\mathrm{B}}T_{\mathrm{L}}\right), \qquad (5)$$
$$P_{\mathrm{d}}=S_{\mathrm{MP}}(I_{\mathrm{d}}/e)+\omega_{\mathrm{e}}. \qquad (6)$$

Therefore, the carrier acceleration integral is eliminated from the considerations. The kinetics of the Si-H bond are described by the system of rate equations [20] and a simplified solution

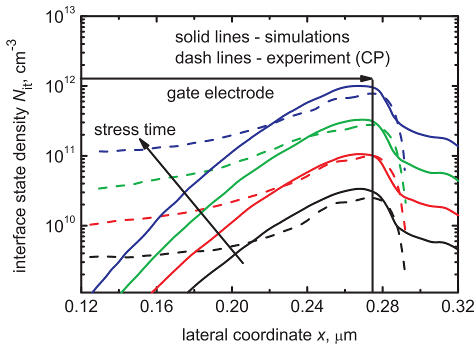

Fig. 12. The N_{it} profiles extracted from charge-pumping data and simulated using the fitting scheme (11) from [51].

of this system leads to a square root time dependence of N_{it}:

$$N_{\mathrm{it}} = (N_0 \lambda_{\mathrm{emi}} [P_{\mathrm{u}}/P_{\mathrm{d}}]^{N_{\mathrm{l}}}) t^{1/2}, \tag{7}$$

where N_0 is the density of the "virgin" bonds.

The Bravaix model distinguishes between three main regimes [50]: The "hot-carrier" regime with low I_{d} but high carrier energies. Here the SP-process plays the dominant role and the "lucky-electron" model is valid, defining the life-time $1/\tau_{\mathrm{SP}} \sim (I_{\mathrm{d}}/W)(I_{\mathrm{s}}/I_{\mathrm{d}})^m$ (I_{s} is the substrate current, W the device width). The opposite case is for high carrier flux with low energies, where the MP-process governs the degradation: $1/\tau_{\mathrm{MP}} \sim [(V_{\mathrm{ds}} - \hbar\omega)^{1/2}(I_{\mathrm{s}}/W)]^{E_{\mathrm{B}}/\hbar\omega} \exp(-E_{\mathrm{emi}}/k_{\mathrm{B}}T_{\mathrm{L}}) \approx [V_{\mathrm{ds}}^{1/2}(I_{\mathrm{d}}/W)]^{E_{\mathrm{B}}/\hbar\omega}$. At moderate V_{ds} and I_{d} the situation is controlled by electron-electron scattering with the corresponding life-time $1/\tau_{\mathrm{EES}} \sim (I_{\mathrm{d}}/W)^2(I_{\mathrm{s}}/I_{\mathrm{d}})^m$. Under real device conditions all the modes are present and a superposition of all the contributions gives the device life-time: $1/\tau_{\mathrm{d}} = K_{\mathrm{SP}}/\tau_{\mathrm{SP}} + K_{\mathrm{EES}}/\tau_{\mathrm{EES}} + K_{\mathrm{MP}}/\tau_{\mathrm{MP}}$.

In the most recent development of the Bravaix model, the SP-process rate is described by the fitting scheme [51]

$$S_{\mathrm{SP}}(E) = 0, \qquad E < 1.5\mathrm{eV} \tag{8}$$

$$S_{\mathrm{SP}}(E) = \mathrm{const}, \qquad 1.5 \le E < 1.9\mathrm{eV} \tag{9}$$

$$S_{\mathrm{SP}}(E) = \alpha \exp(3E), \qquad 1.9 \le E < 2.5\mathrm{eV} \tag{10}$$

$$S_{\mathrm{SP}}(E) = \beta (E - 1.5)^{11}, \qquad E \ge 2.5\mathrm{eV} \tag{11}$$

to better fit the charge-pumping data. Using the above, the authors have managed to eliminate the carrier transport problem and the information about the DF, while obtaining good agreement between N_{it} profiles extracted from charge-pumping data and simulated ones, Fig. 12.

The main disadvantage of the model is the missing carrier transport treatment which allows one to distinguish between the SP-mechanism, electron-electron scattering, and MP-process driven modes. As will be shown, all these mechanisms have to be considered consistently because they affect each other via the carrier DF, and hence control HCD. As an adjacent problem, the proposed scheme for the SP-process rate is based on fitting parameters, not on physical mechanisms.

D. A Detailed Model Including Carrier Transport

To overcome the above mentioned limitation, we have proposed and verified a more detailed approach for HCD

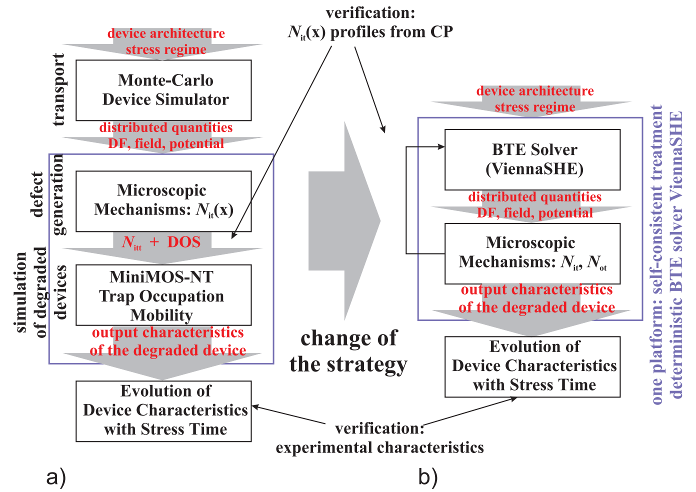

Fig. 13. Our HCD model includes the carrier transport module, a module describing the defect generation, and a module responsible for modeling of the degraded devices (a). Different subtask of the problem are linked but not considered self-consistently. The platform for self-consistent consideration is a deterministic Boltzmann transport equation solver ViennaSHE (b).

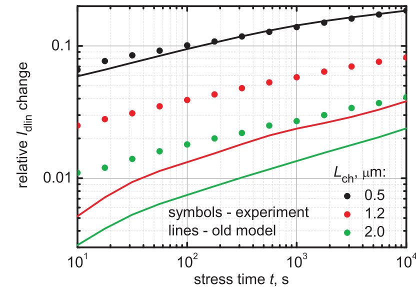

Fig. 14. The model calibrated for the 0.5μm transistor [36] fails while trying to reproduce I_{dlin} degradation in 1,2 and 2,0μm devices.

modeling which tries to more accurately capture the physical picture behind this phenomenon [52], [53]. We aim at covering and linking all the levels related to this effect, starting from the microscopic mechanisms of defect generation and ending at the device level. A physics-based model of HCD may be conditionally separated into three main sub-tasks: the carrier transport module, a module describing the defect build-up, and a module responsible for the simulation of the degraded devices [52], [53]. The flowchart of the model is sketched in Fig. 13a. Carrier transport is treated with a full-band Monte-Carlo device simulator, which calculates the carrier DF for a particular device architecture and conditions. The information on the DF is used to generate N_{it} profiles, which are then loaded to a circuit and device simulator MINIMOS-NT [54], which finally calculates the characteristics of the degraded device. At the moment we solve these three subtasks subsequently. A self-consistent solution of the problem can be achieved if all these modules are integrated within the same platform, Fig. 13b. For this purpose a deterministic Boltzmann transport equation solver ViennaSHE developed in our institute will be used. A first implementation of our HCD model incorporated into ViennaSHE was already demonstrated in [55].

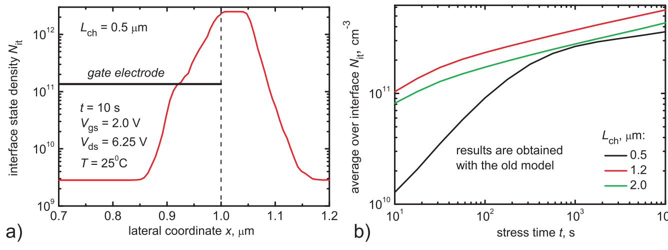

Fig. 15. The N_{it} distribution along the interface calculated with the first model [36] considering only the contribution to the damage provided by minority carriers (a) and average N_{it} concentration calculated with the first model for 0.5, 1.2, and 2.0μm devices.

The main requirement to the model is that it has to be able to represent HCD observed in different MOSFETs *using a single set of model parameters*. The first version of the model considered only the contribution of the minority carriers, for instance electrons in an n-MOSFET. The model was able to successfully represent the I_{dlin} degradation in a wide range of stress/operating conditions in the case of a 5V n-MOSFET with a channel length 0.5 μm, see [36], [56].

To further verify the model, a series of three 5V n-MOSFETs with identical architecture differing only in channel lengths (L_{ch} = 0.5, 1.2, and 2.0 μm) was used [52]. Devices were subjected to a hot-carrier stress at V_{gs} = 2.0V and V_{ds} = 6.25V. The model was calibrated in a manner to represent the I_{dlin} change vs. time. Quite strikingly, the first version of the model completely fails while trying to capture I_{dlin} degradation in MOSFETs with different channel lengths with the same set of parameters. Fig. 14 shows that for devices with L_{ch} of 1.2 and 2.0μm, $\Delta I_{\mathrm{dlin}}(t)$ predicted by the model is severely underestimated.

The reason is that the concentration of interface states generated by primary channel electrons peaks outside the channel (see Fig. 15a), thereby only weakly affecting the device performance. Fig. 15b shows the average N_{it}, that is $N_{\mathrm{it}}(x)$ integrated over the interface and divided by the interface length. One can see that more severe degradation (0.5 μm device in Fig. 15a) corresponds to a lower average concentration of interface states, see Fig. 15b. This suggests that longer devices should be less less sensitive to interface traps generated by primary channel electrons.

Thus, another mechanism leading to interface states created in the channel has to be responsible for this discrepancy. The most probable candidate is the secondary carriers generated by impact ionization (holes in this case). Indeed, when generated, these holes are accelerated by the electric field towards the source and thus create interface states inside the channel. Therefore, secondary generated majority holes cannot be ignored and for a proper HCD treatment one has to consider both types of carriers.

In the improved version of the model, the carrier transport module evaluates the DFs for both electrons and holes at each point at the interface. Both SP- and MP-mechanisms are controlled by the carrier acceleration integral:

$$I_{=}\int_{E_{\mathrm{th}}}^{\infty} f(E)g(E)\sigma(E)v(E)\mathrm{d}E \tag{12}$$

where $f(E)$ is the carrier DF, $g(E)$ the density-of-states, v the carrier velocity, and $\sigma(E)$ is the reaction cross section. Expressions for electrons and holes as well as for SP- and MP-processes have the same functional form. For the SP-process, the superposition of electron and hole AIs weighted with the attempt frequencies $\nu_{\mathrm{SP,e}}$, $\nu_{\mathrm{SP,h}}$ gives the generation rate for this mechanism. Assuming first-order kinetics, one obtains this dependence of interface trap density vs. time for the SP-process:

$$N_{\mathrm{SP}}(t)=N_0\left[1-\mathrm{e}^{-(\nu_{\mathrm{SC,e}}I_{\mathrm{SC,e}}+\nu_{\mathrm{SC,h}}I_{\mathrm{SC,h}})t}\right]. \tag{13}$$

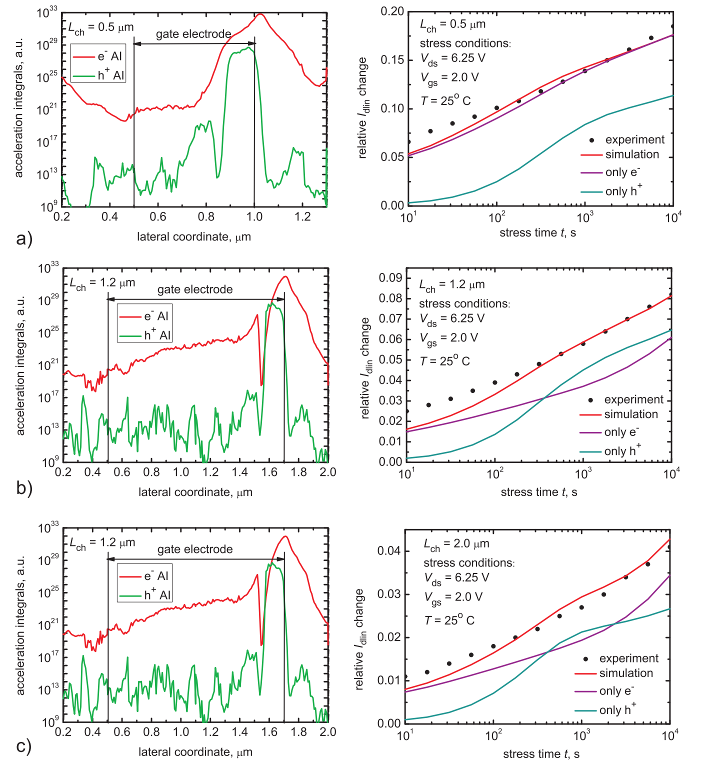

Fig. 16. The AIs for electrons and holes (left panel) and the linear drain current degradation ΔI_{dlin} (right plots) for 0.5 (a), 1.02 (b), and 2.0μm (c) MOSFETs. The right plots show experimental vs. calculated $\Delta I_{\mathrm{dlin}}(t)$ as well as contributions provided by exclusively electrons and holes.

As for the MP-process, a Si-H bond is treated as a truncated harmonic oscillator and the solution of the system of rate equations is [36], [52]:

$$N_{\mathrm{MC}}=N_0\left(\frac{\lambda_{\mathrm{emi}}}{P_{\mathrm{pass}}}\left(\frac{P_{\mathrm{u}}}{P_{\mathrm{d}}}\right)^{N_{\mathrm{l}}}\left(1-\mathrm{e}^{\lambda_{\mathrm{MC}}t}\right)\right)^{1/2}. \tag{14}$$

As the secondary holes are generated by impact ionization and are then accelerated by the electric field towards the source, the thereby created interface states are shifted with respect to

the electron-induced ones. Since in our model the acceleration integral is the driving force of HCD, the AI clearly reflects this tendency. Indeed, the hole AIs feature maxima which are shifted towards the source as compared to the maxima of the electron AIs, see Fig. 16, left panels. Finally, the model is able to capture $\Delta I_{\mathrm{dlin}}(t)$ observed in three different devices with a single set of parameters, see Fig. 16, right plots.

The lateral profile of the interface state density $N_{\mathrm{it}}(x)$ and the density of the bulk oxide traps $N_{\mathrm{ot}}(x)$ is plotted in Fig. 17. These profiles were extracted from charge-pumping data using the techniques presented in [57]. The $N_{\mathrm{it}}(x)$ profiles feature two peaks. These peaks are related to the contributions induced by primary channel electrons and secondary generated holes and correspond to the maxima of the electron and hole AIs [57]. The second important message is that the extraction procedure demonstrates that bulk oxide traps do contribute to HCD.

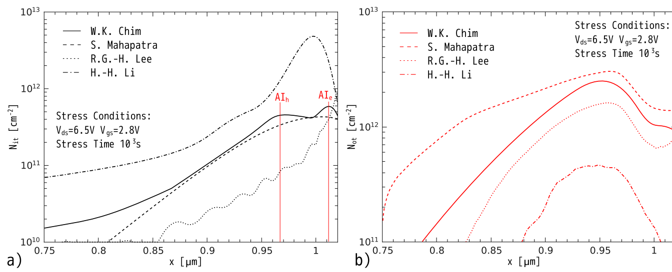

Fig. 17. The interface state density N_{it} (a) and the bulk oxide trap density N_{ot} (b) profiles extracted from charge-pumping data [57]. One can see that the N_{it} profiles have two maxima related to primary channel electrons and secondary generated holes. These maxima correspond to the peaks of electron and hole acceleration integrals.

V. Open issues in understanding of HCD

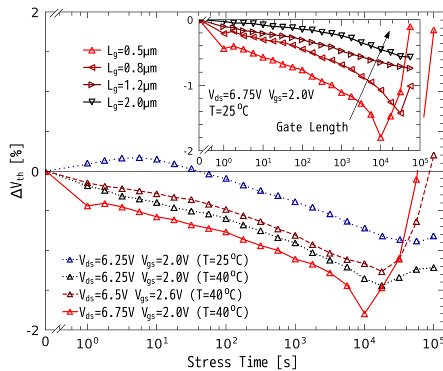

Fig. 18. The turn-around effect of the threshold voltage shift. First V_{th} decreased due to the hole trapping in the oxide bulk. After ~10ks this mechanism was overtaken by the interface electron trapping.

A still open question is to what extend bulk oxide traps contribute to hot-carrier degradation, as there are several arguments against and in favor. First, we have recently discussed a turn-around of the V_{th} shift [5], see Fig. 18. The device was stressed close to the worst-case conditions and the threshold voltage was monitored up to 100ks. Initially V_{th} decreases due to hole trapping in the oxide bulk while after 10ks it starts to increase due to trapping of electrons by interface traps. At approximately 100ks the second effect has overcompensated the hole trapping in the oxide bulk. At the presence of only interface charges the charge-pumping current plotted as a function of the high level of the gate pulse (for the varying high-level technique) would have saturated, resulting in a plateau. This is not the case in Fig. 19, suggesting the contribution of bulk oxide traps.

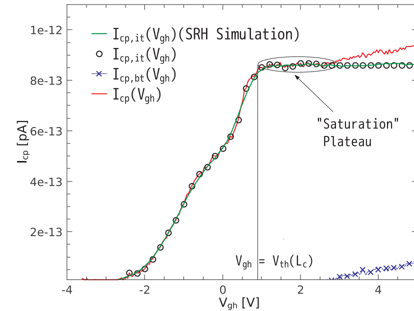

Fig. 19. The charge-pumping current as a function of the high level of the gate pulse. The signal does not have a plateau but continues to increase, thereby suggesting a bulk oxide trap contribution.

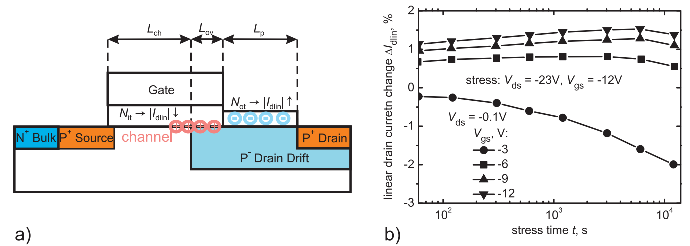

Fig. 20. A schematic representation of the DEMOS with electrons and holes trapped in different sections of the device (a). The turn-around of the linear drain current (b). At short times, I_{dlin} increases due to the electron trapping, while hole trapping dominates the I_{dlin} decrease at long stress times.

The linear drain current turn-around effect was also observed previously, see [4]. A p-channel DEMOS transistor (Fig. 20a) was stressed at V_{ds}=-23V and V_{gs}=-12V. I_{dlin} first increases and then decreases, see Fig. 20b. This is because charges of opposite signs are trapped in different sections of the transistor, Fig. 20a. For short stress times the I_{dlin} change is dominated by electron trapping in the drift region outside the poly-gate. Vice versa, hole trapping in the drift region under the poly-gate governs I_{dlin} degradation at long stress times

At the same time BTI typically has a pronounced recoverable component, and this component is related to switching oxide traps [37]–[39]. Reasoning by analogy, one concludes that if trapping/detrapping in the oxide bulk contributes to HCD, a recoverable component should be present. But only high-voltage devices (LDMOS, DEMOS) have recovery while CMOS transistors usually do not show any recovery at all. To summarize, the puzzle whether bulk oxide traps contribute to hot-carrier degradation is still not reconciled and requires more investigation.

VI. Conclusion

Hot-carrier degradation is related to the generation of interface and most probably bulk oxide defects during stress. The

generation is triggered by the bombardment of the interface by carriers which were accelerated by the electric field. HCD is strongly localized phenomenon. Traditionally HCD has been associated with "hot" carriers which can trigger the Si-H bond dissociation in a single collision (the single-particle process). However, this has changed in ultra-scaled devices when several "colder" carriers can trigger the bond dissociation (multiple-carrier process). The interplay between these SP- and MP-mechanisms results in different worst-case conditions of HCD in long- and short-channel MOSFETs. Also, in ultra-scaled devices HCD has a different temperature dependence, that is the degradation becomes more severe at higher temperatures, while in long-channel devices HCD is suppressed when the temperature is increased.

The change of the temperature dependence is due to electron-electron scattering, which plays a dominant role in scaled devices. Even more, HCD is controlled by the carrier distribution function and for a proper description of this phenomenon one has to carefully solve the carrier transport problem. The information on the carrier distribution thus defines the defect generation kinetics. Finally, generated defects can affect the device electrostatics and act as additional scattering centers, thereby degrading the mobility. To predict the device life-time, a thorough modeling of the device characteristics is also important.

We have analyzed the most important physics-based models of HCD highlighting their limits of validity. Finally, we presented our own modeling attempt which incorporates contributions to the damage provided by both types of carriers. The model is able to represent HCD in transistors with different channel-lengths using a single set of model parameters.

Some issues in understanding and modeling of hot-carrier degradation still remain open. The most important question is whether bulk oxide traps contribute to HCD or not. The main arguments in favor is given by the turn-around effects related to trapping of different charges by interface traps and bulk oxide states. At the same time, BTI has a prominent recoverable component related to switching oxide traps. However, HCD shows recovery only in high-voltage devices while the damage in scaled MOSFETs remains permanent. This issue is the subject of the further research.

ACKNOWLEDGMENT

The authors acknowledge support by the Austrian Science Fund (FWF), grant P23598.

REFERENCES

[1] A. Goetzberger, IEEE Proceedings **54**, 1454 (1966).
[2] M. Ancona *et al.*, IEEE Trans Electron Dev. **35**, 221 (1988).
[3] P. Heremans *et al.*, IEEE Trans. Electron Dev. **35**, 2194 (1988).
[4] J. Chen *et al.*, Jpn. Journ. Appl. Phys. **48**, 04C039 (2009).
[5] I. Starkov *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2012), pp. 1–6.
[6] C. Helms *et al.*, Reports on Progress in Physics **57**, 791 (1994).
[7] C. Hu *et al.*, IEEE Trans. Electron Dev. **48**, 375 (1985).
[8] D. DiMaria *et al.*, Journ. Appl. Phys. **65**, 2342 (1989).
[9] D. DiMaria *et al.*, Journ. Appl. Phys. **89**, 5015 (2001).
[10] A. Bravaix *et al.*, *Proc. European Symposium on Reliability of Electron Devices Failure Physics and Analysis (ESREF), tutorial* (2010).
[11] S. Rauch *et al.*, *Proc. International Reliability Physics Symposium (IRPS), tutorial* (2010).
[12] W. McMahon *et al.*, IEEE Trans. Nanotech. **2**, 33 (2003).
[13] B. Ricco *et al.*, *Proc. International Electron Devices Meeting (IEDM)* (84), pp. 92–95.
[14] T. Mizuno *et al.*, *Proc. International Electron Devices Meeting (IEDM)* (1992), pp. 695–698.
[15] J. Bude, *Proc. VLSI Symposium Tech. Digest* (1995), pp. 101–102.
[16] F.-C. Hsu *et al.*, IEEE Electron Dev. Lett. **5**, 148 (1984).
[17] K. Hess *et al.*, Circuits and Devices Mag. 33 (2001).
[18] A. Haggag *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2001), pp. 271–279.
[19] W. McMahon *et al.*, *Proc. Int. Conf. Mod. Sim. Micro* (2002), Vol. 1, pp. 576–579.
[20] A. Bravaix *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2009), pp. 531–546.
[21] M. Song *et al.*, IEEE Trans Electron Dev. **44**, 268 (1997).
[22] J. Wang-Ratkovic *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2003), pp. 312–314.
[23] W. Qin *et al.*, Semicond. Sci. Technol. **13**, 453 (1998).
[24] V. Reddy, *Proc. International Reliability Physics Symposium (IRPS), tutorial* (2004).
[25] E. Li *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (1999), pp. 253–258.
[26] C. Lin *et al.*, *Proc. International Electron Devices Meeting (IEDM)* (2000), pp. 135–138.
[27] J. Lyding *et al.*, Appl. Phys. Lett. **68**, 2526 (1996).
[28] K. Lee *et al.*, IEEE Electron Dev. Lett. **29**, 389 (2008).
[29] M. Jo *et al.*, Appl. Phys. Lett. **94**, 053505 (2009).
[30] E. Amat *et al.*, Microel. Engineering **87**, 47 (2010).
[31] P. Childs *et al.*, IEEE Electron Dev. Lett. **31**, 139 (1995).
[32] S. Rauch *et al.*, IEEE Electron Dev. Lett. **19**, 463 (1998).
[33] A. Ghetti *et al.*, IEEE Trans Electron Dev. **46**, 696 (1999).
[34] S. Rauch *et al.*, IEEE Trans Dev. Material. Reliab. **1**, 113 (2001).
[35] A. Abramo *et al.*, IEDM – Techn. Digest 301 (1995).
[36] S. Tyaginov *et al.*, *Proc. International Symposium on the Physical & Failure Analysis of Integrated Circuits (IPFA)* (2010).
[37] V. Huard *et al.*, Microel. Reliab. **46**, 1 (2006).
[38] T. Grasser *et al.*, IEEE Trans Electron Dev. **58**, 3652 (2011).
[39] T. Grasser, Microel. Reliab. **52**, 39 (2012).
[40] B. Tuttle *et al.*, Phys. Rev. B **59**, 12884 (1999).
[41] R. Walkup *et al.*, Phys. Rev. B **48**, 1858 (1993).
[42] B. Persson *et al.*, Surface Science **390**, 45 (1997).
[43] J. Lyding *et al.*, Appl. Surf. Sci. **13-132**, 221 (1998).
[44] K. Stokbro *et al.*, Phys. Rev. Lett. **80**, 2618 (1998).
[45] R. Biswas *et al.*, Appl. Phys. Lett. **72**, 3500 (1998).
[46] K. Hess *et al.*, Physica E **3**, 1 (1998).
[47] S. Rauch *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2005).
[48] C. Guerin *et al.*, Journ. Appl. Phys. **105**, (2009).
[49] A. Bravaix *et al.*, *Proc. International Reliability Physics Symposium (IRPS)* (2010), pp. 55–64.
[50] C. Guerin *et al.*, IEEE Trans. Dev. Material. Reliab. **7**, 225 (2007).
[51] Y. Randriamihaja *et al.*, Microel. Reliab. **52**, 2513 (2012).
[52] S. Tyaginov *et al.*, *Proc. International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (2011), pp. 123–126.
[53] S. Tyaginov *et al.*, *Proc. European Solid-State Device Research Conference (ESSDERC)* (2011), pp. 151–154.
[54] *MiniMOS-NT Device and Circuit Simulator*, Institute for Microelectronic, TU Wien.
[55] M. Bina *et al.*, *Proc. International Electron Devices Meeting (IEDM)* (2012), p. accepted.
[56] S. Tyaginov *et al.*, Microelectronics Reliability **50**, 1267 (2010).
[57] I. Starkov *et al.*, *Proc. International Symposium on the Physical & Failure Analysis of Integrated Circuits (IPFA)* (2012), pp. 1–6.