We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Transistor Degradations in Very Large-Scale-Integrated CMOS Technologies

Chang Yeol Lee

Additional information is available at the end of the chapter

Abstract

The historical evolution of hot carrier degradation mechanisms and their physical models are reviewed and an energy-driven hot carrier aging model is verified that can reproduce 62-nm-gate-long hot carrier degradation of transistors through consistent aging-parameter extractions for circuit simulation. A long-term hot carrier-resistant circuit design can be realized via optimal driver strength controls. The central role of the V_{GS} ratio is emphasized during practical case studies on CMOS inverter chains and a dynamic random access memory (DRAM) word-line circuit. Negative bias temperature instability (NBTI) mechanisms are also reviewed and implemented in a hydrogen reaction-diffusion (R-D) framework. The R-D simulation reproduces time-dependent NBTI degradations interpreted into interface trap generation, ΔN_{it} with a proper power-law dependency on time. The experimental evidence of pre-existing hydrogen-induced Si–H bond breakage is also proven by the quantifying R-D simulation. From this analysis, a low-pressure end-of-line (EOL) anneal can reduce the saturation level of NBTI degradation, which is believed to be caused by the outward diffusion of hydrogen from the gate regions and therefore prevents further breakage of Si–H bonds in the silicon-oxide interfaces.

Keywords: hot carrier injection (HCI), hot carrier degradation (HCD), hot carrier-resistant design, negative bias temperature instability (NBTI), reaction-diffusion (R-D) of hydrogen

1. Introduction

Since the concept of an integrated circuit was first proposed by Jack Kilby in 1958, and a first version of a self-aligned poly-silicon gate CMOS-integrated circuit was fabricated at Fairchild® in 1968, integrated circuit technology has led to unprecedented thriving and prosperity in the electronic industry for the last half century. The initial integration of a number of transistors started with only a few tens in a circuit, which we call small-scale integration (SSI) has today

expanded to a few billions, called very large-scale integration (VLSI) or ultra large-scale integration (ULSI). The annual growth rate of the number of transistors per IC has followed a well-known formula, Moore's Law, which indicates that the density of devices per chip doubles every 18 months, that is, the population of transistors in a chip increases by 1000 times every 15 years. Although in the beginning it was a merely an observation, it evolved into a de facto mandatory target for cutting-edge technology developers. For example, the Intel® CPU transistor count has faithfully followed Moore's Law for four decades (1971–2012) (**Figure 1**). Specifically, their newest microprocessor, Ivy Bridge Core i7™, possesses 1,400,000,000 transistors [1]. It is a 609,000× increase in transistor count from its 1971 version. Such an exponential increase in integration number is attributed to a series of successes in shrinking feature size. The benefit of scaling is obvious: more integrated transistors enable more sophisticated data-driven operations and less switching delays per logic gate, thereby enhancing data transaction bandwidth. More data with enhanced speed play a decisive role in the rapid growth of the information and communication industry.

During the continuous pursuit of scaling, the following inherent issues have arisen:

- The challenge to sustain photolithographic pattern fidelity and critical dimension (CD) uniformity becomes profound as dimension scaling and integration levels increase.

- As transistor gate length is shrunk, electric field strength inside of a transistor increases and more degradation may occur in devices. Vertical and lateral e-fields can be mitigated by reduction of bias voltage, V_{DD}. Its minimum level tends to be limited by the minimum threshold voltage of logic gates defined by distinguishable high states against thermal noise and permissible off-leakage currents. Since the V_{DD} limit of around 1.0 V is already achieved in state-of-the-art technologies, inevitable increases of internal e-fields inside of devices may define the practical limit of technology scaling.
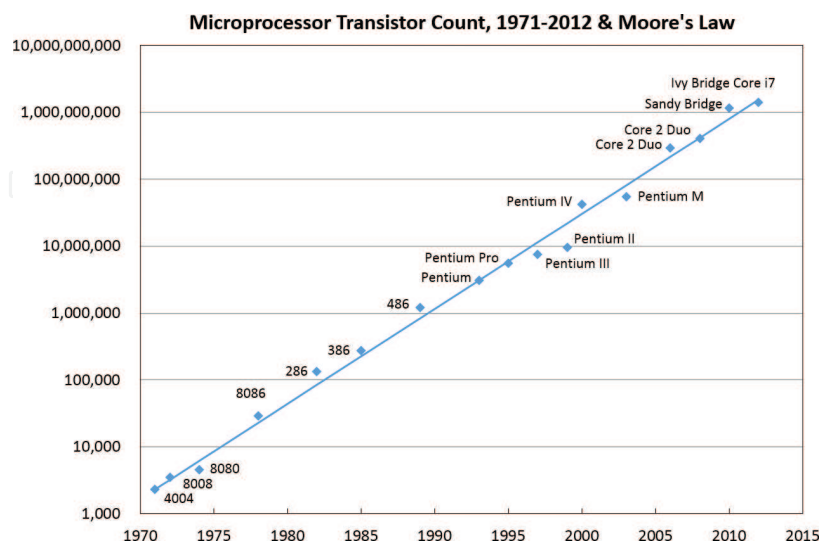

Figure 1. Intel® microprocessor transistor count evolution between 1971 and 2012. For the last 40 years, it has consistently behaved as predicted by Moore's Law. (Source: Intel Corp, Web Page [1]. Figure 1 is a graphical translated version, which was originally expressed in a tabular form in [1].)

The first issue is directly related to how short are the wavelengths of photolithographic light sources we employ. A 193-nm ArF light source with immersion ambience is known to have 40-nm patterning capability as its best performance. Since a feasible solution for a new light source with a lesser wavelength than ArF has not yet been found, a complicated combination of photolithography and etching processes, for example, a double-spacer pattern technology (D-SPT), is employed to enable the latest 10-nm range patterns. Such a complicated combination of critical steps may cause a sizable variation of CDs. As a consequence, much expertise and extensive trials and corrections are required to achieve pattern optimization, which becomes the confidential property of cutting-edge companies.

So solving the first issue relies in large part on skill and trial-and-error correction processes. By contrast, the second issue is related largely to scientific analysis. When bias is applied to a scaled MOSFET, a localized e-field is established in the drain side, which can accelerate mobile carriers (electrons or holes) passing this region. Some of the accelerated carriers can trigger an avalanche multiplication process, which increases the possibility of generating energetic carriers that can surmount the energy barrier between silicon/silicon dioxide or cause damage in Si/SiO_2 interface. Energetic carriers or "hot carriers" can also be generated by energy exchange during carrier-carrier-scattering processes. This kind of device degradation mechanism is called "hot carrier injection (HCI)," or "hot carrier degradation (HCD)" and is regarded as a typical degradation mechanism driven by high lateral e-fields or V_{DD}. The most efficient prevention of this kind of degradation is reducing V_{DD}. Large efforts have been devoted in device and circuit research to develop power-efficient and degradation-aware low V_{DD} transistors and circuit solutions. Despite this effort, there still inevitably remain high voltage needs in some specific applications, like word-line decoders in dynamic random access memory (DRAM) circuits.

Dynamic Random Access Memory (DRAM) is one of the most popular memory devices featuring high data read/write speed with low bit cost. Compact placement of a single-bit storage capacitor and its switch transistor composes a DRAM cell. To avoid large off-leakage caused by high e-fields and thereby insufficient data retention capability, a three-dimensional (3-D) recess-channel scheme has recently been developed to reduce the e-field strength by extending the channel length. Although sufficient data retention time can be achieved by the 3-D recess-channel structures, reduction of the channel conductance of the long channel length expensively undermines the access speed. Non-scaled gate voltage can compensate for this loss. Such a decoupling from scaling rules (planar dimensions scale down, whereas gate voltages do not) may cause HCD issues in cell gate bias-pumping voltage (V_{PP}) circuits. Since HCD is ascribed to the high electric field and/or high gate voltage, the mitigation strategy largely relies not only on device internal structure and doping profiles but also on the circuit and layout strategy. More detailed descriptions based on practical case studies and some general guidelines for the HCI-resistant circuit design can be found in the next section.

Lateral (channel length) shrinkage should be indispensably coupled with vertical (gate oxide thickness) shrinkage to maintain "long channel-like" transistor characteristics. The key enabler of vertical scaling is the superb electrical and material properties of silicon dioxide. As to silicon dioxide, only about 10 stacks of molecules can provide good isolation under 5.5–6.0 MV/cm of

electric field intensity or can sustain "off-characteristics" of the few tens of nanometer-scaled MOSFET. Despite its stability, modern plasma-intensive fabrication processes can induce multiple charging in the gate electrodes, which can generate a number of silicon-oxide bond breakages. Most bond breakages can be passivated by end-of-line (EOL) hydrogen or deuterium passivation steps in order to electrically deactivate the dangling bonds. In this circumstance, another kind of device degradation mechanism can be triggered: negative bias temperature instability (NBTI) can be activated by moderated gate bias and temperatures applied in p-channels where abundant inverted holes and hydrogen-passivated silicon-dangling bonds exist. Although its mechanism is still not completely understood, atomic and/or molecular hydrogen reactions and diffusions associated with passivated Si/SiO$_2$ interfaces are widely accepted to define how and how much degradation takes place. In Section 3, a quantitative analysis is provided based on the reaction-diffusion of hydrogen simulation. A mitigation strategy for the long-term NBTI degradation is also suggested during the analysis.

In this chapter, studies on the most typical scaled-down-related device reliability issues, HCD in NMOS and NBTI in PMOS, are presented with practical case studies to attempt to broaden the reader's knowledge of device degradation and its impact on advanced CMOS scaling.

2. Hot carrier degradation

2.1. Historical review

Hot carrier degradation (HCD) is one of the typical wear-out degradation mechanisms that causes catastrophic failures in systems. This kind of failure may implosively trigger irreversible and unrecoverable damage in systems. Readers can find typical cases of HCD failure syndromes and their impacts on complete products in Ref [2]. Many investigations have been conducted to reveal the transistor degrading hot carrier generation mechanism. The first successful theory structure was announced as the lucky electron model (LEM) suggested by Hu et al. in 1985 [3]. The LEM is regarded as a classic theory and has been widely used so far because it allows the depiction of a clear image for hot carrier generation and its role in creating interface traps. It focuses on e-field-driven hot carrier generation. A quasi two-dimensional analysis of Poisson's equation derives the exponential shape of e-fields in the velocity saturation region (VSR) and a sharp peak of e-fields built in front of the neutral drain region [4]. All the energy gain processes are assumed to be concentrated in the peak e-field spot where lucky electrons are generated by re-directional impact ionizations with Si-lattices. As a result, lucky electrons can surmount the Si/SiO$_2$ energy barrier (3.2 eV) and generate the Si interfacial traps, N_{it} [3]. On this basis, the following interface generation rate, r_{it}, was derived:

$$
\begin{aligned}
r_{it,LEM} &= C_1 \cdot I_D \cdot \left(\frac{I_B}{I_D}\right)^{\Phi_{it}/\Phi_{ii}} \\
&= C_1 \cdot I_D \cdot \left[A \cdot (V_D - V_{DSAT}) \cdot \exp\left(\frac{-B\ell}{V_D - V_{DSAT}}\right)\right]^{\Phi_{it}/\Phi_{ii}}
\end{aligned}
\tag{1}
$$

where I_B and I_D are substrate current and drain current, respectively, Φ_{ii} is the impact ionization threshold energy (1.3 eV for electrons) and Φ_{it} is the interface state generation threshold

energy (3.7 eV for electrons), ℓ is the characteristic length of VSR, and A, B, and C are constants, respectively. The power-law exponent, $\Phi_{\mathrm{it}}/\Phi_{\mathrm{ii}}$, is calculated to be 2.8 and this approximately matches experimental results that guarantee that the formula captures the correct image of HCD.

Transistor degradations mean that threshold voltages shift, mobility decreases, and drain-extrinsic resistances increase, all of which are ascribed to the interface trap generation. An isotope effect found by the scanning tunneling microscope (STM) method [5] and HCD experiments using hydrogen and deuterium-annealed samples [6] reveals that the dissociation of hydrogen from the interfacial Si–H bonds by injected energetic electrons can lead to unrecoverable degradations. The interface trap generation rate is empirically expressed as

$$\Delta N_{\mathrm{it}}(t) \propto (r_{\mathrm{it}} \cdot t)^{\,n} \tag{2}$$

where n is a time exponent that has been known to be around 0.5, which can be derived from a hydrogen diffusion-limited process [3]. An assumption of high diffusivities of hydrogen in silicon dioxide and in polysilicon gate regions is required to describe the 0.5 dependency. More specifically quickly removing the hydrogen from the interface and therefore also the repassivation process cannot dominantly influence the whole hydrogen reaction-diffusion process. Contradictable findings have been also reported in PMOS-negative bias temperature instability (NBTI) research [7]. Fast-diffused hydrogen in the SiO_2 region slows down in the silicon-nitride interface and in the polysilicon region due to small diffusion constants in those regions. This results in an accumulation of hydrogen in the SiO_2 region, which strikes a balance between dissociation and repassivation of silicon-dangling bonds. As a consequence, the interface state generation rate decreases to produce a smaller n (1/4–1/6). The discrepancy of time exponents between NMOS HCI and PMOS NBTI can be ascribed to the difference in the stressed area (only localized to the peak e-field spot in HCI vs. the whole gate oxide area in NBTI) and its influence on hydrogen diffusion profiles: an increasingly wider diffusion front of hydrogen in NMOS HCD enhances the ΔN_{it} rate more than that of PMOS NBTI where consistent one-dimensional diffusion of hydrogen occurs [8]. Furthermore, asymmetric behaviors between NMOS and PMOS (a large amount of degradation is quickly recovered when the stress biases are removed in PMOS, while no substantial recovery takes place in NMOS) imply the different nature of the Si-H dissociation produced by cold holes injected during NBTI stress and by hot electrons injected during HCI stress. The existence of deep-level hole traps (DLHT) [9, 10] was proposed to draw a plausible picture of the asymmetric behavior. In the author's opinion, more studies are still needed to reveal the underlying physics for a comprehensive understanding.

As the e-field-driven (and consequently the applied voltage-driven) LEM reveals the HCD mechanism to be successful, it also instigates a voltage scale-down from 3.3–5.0 to 1.8–2.5 V in the shrunk gate length transistors in an effort to avoid HCD risk. However, HCD still remains against expectation in the 1.8–2.5-V regime. Neither hot electron injections into gate oxide (requires 3.2 eV at least) nor interface trap generations (requires 3.7 eV) may take place according to LEM because of insufficient driving voltage. A new hypothesis for the HCI generation mechanism, electron-electron scattering (EES), has been proposed to explain the

hot carrier generation in medium V_{DD} conditions. This hypothesis has been accepted through numerous experimental verifications [11]. It involves an energy-exchanging electron-scattering process to generate hot electrons under moderate bias conditions. A doubling of its energy can be achieved when a perfect elastic collision between the excited electrons, 1.8–2.5 V, is sufficient to generate the interface-degrading hot electrons. The mathematical expression of EES implies these aspects as follows:

$$r_{\text{it,EES}} = C_2 \cdot I_D{}^2 \cdot \left(\frac{I_B}{I_D}\right)^m.$$ (3)

Note that the power-law exponent of I_D changes from 1 as in Eq. (1) to 2 as in Eq. (3) reflecting a statistical interaction of two independent sources for EES. This secondly found hot carrier generation mechanism dominates in the sub-micrometer range-scaled MOSFETs whose drain currents have a range between 40 and 500 µA/µm with high V_{GS} drive [11, 12].

Further voltage scale-down to 1.0–1.2 V might extinguish any possibility of hot carrier generation via LEM or EES mechanisms. A newly developed Si–H bond breakage model has been proposed and demonstrated in deca-nanometer-scaled transistors [13]. Multivibrational hydrogen release (MVHR) is the third kind of mechanism, which is activated through high current injection (the minimum threshold is known to be 1.5 mA/µm), with weak voltage dependency. Since an electron can transfer its kinetic energy to the silicon lattice via optical phonon resonance, the multiple striking of electrons into Si–H bond can lead hydrogen to multiple jumps in its energy state to approach the bond-breaking threshold energy, E_B. This kind of hydrogen-dissociated process, namely multiple vibrational hydrogen release (MVHR), can be triggered by low-energy cold carriers in a sub-1-V-biased channel [13]. The quantum mechanical picture of the process is illustrated in **Figure 2**, and the mathematical expression was proposed to fit the experimental data as

$$r_{\text{it,MVP}} = C_3 \cdot \left[V_{DS}{}^{1/2}\left(\frac{I_D}{W}\right)\right]^{E_B/\hbar\omega} \cdot \exp\left(\frac{-E_{\text{emi}}}{k_B T}\right)$$ (4)

where E_{emi} (~0.26 eV) is the barrier height of E_B from the highest energy state of bonded hydrogen. The unit resonance energy per single phonon excitation, $\hbar\omega$ (~0.075 eV), and threshold energy, E_B (~1.5 eV), defines the required number of the phonon excitation to be $\frac{E_B}{\hbar\omega} = 20$. Due to its tremendously strong power-law dependency on the drain current, a special caution should be paid not to let the transistor's drain current exceed the threshold of the third kind of HCD in any type of transistor operations including burn-in tests; otherwise, very quick wear-out failures may take place. The three kinds of hot carrier generation mechanisms are illustrated in **Figure 3** compared with a large set of experimental data.

To summarize the history of HCD mechanism finding, LEM dominates when $V_{DS} \geq 3.0$ V. Energy-driven or current-driven multiple-particle (MP) mechanisms, EES for 40–500-µA/µm driving range and MVHR for even higher ranges, are subsequently developed. The maximum applied voltage and the minimum duty cycle of CMOS logic design guidelines have been
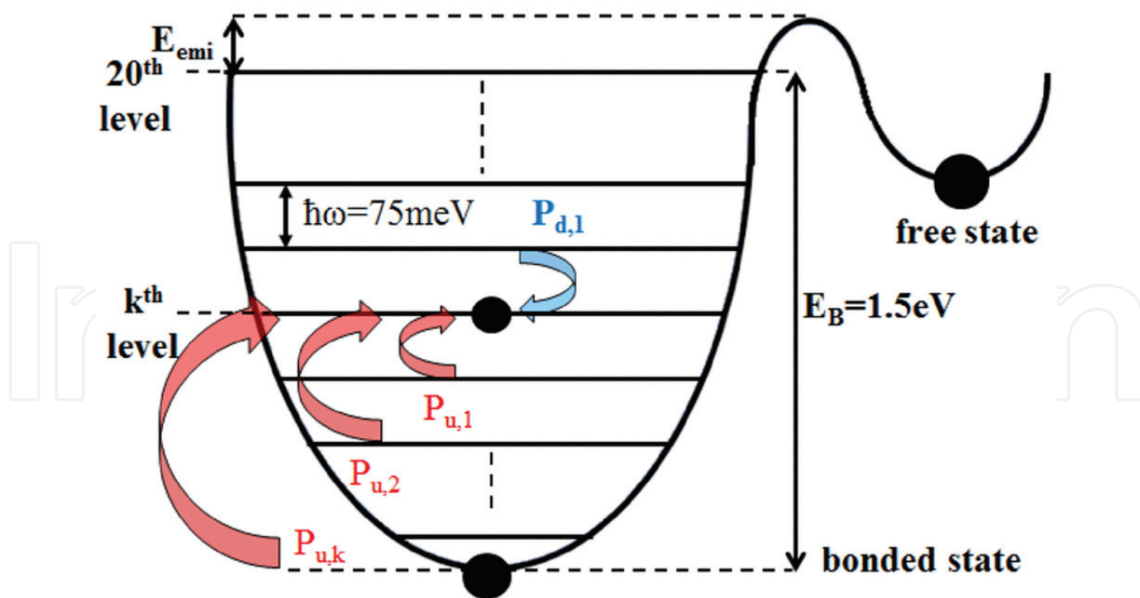
Figure 2. Schematic of Si-H potential well showing the processes increasing the occupancy of the kth level, coming from direct excitation, giving any number of energy quanta between 1 and k, and from the de-excitation of the $(k+1)^{th}$ level. $P_{u,i}$ and $P_{d,i}$ are, respectively, the probability of excitation and de-excitation, giving or losing i energy quanta. Reprinted with permission from Ref. [11], © 2013 IEEE.
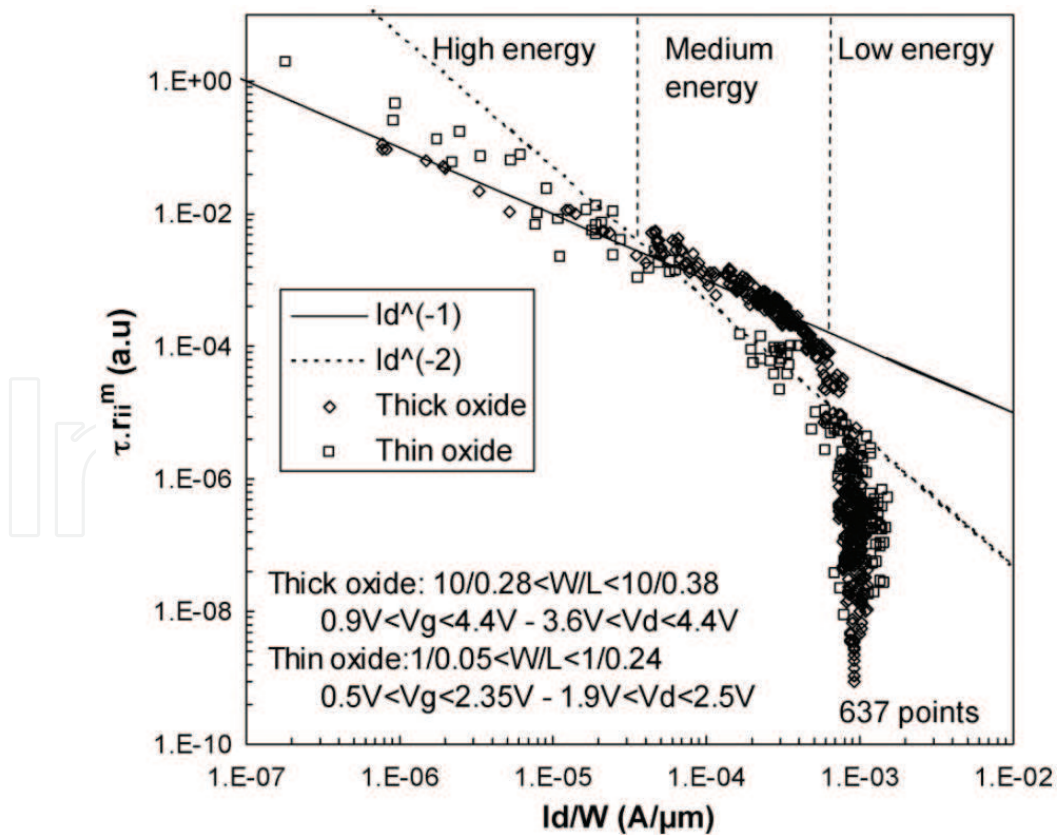

Figure 3. $\tau \cdot r_{ii}^m$ normalized lifetimes versus drain current showing three regimes of hot carriers. Reprinted with permission from Ref. [14], © 2007 IEEE.

made in strong awareness of the e-field-driven HCD. The maximum current-limiting constraints to prevent the current-driven HCD have not yet been made according to the author's knowledge. It might not be required since the ultimately scaled 3.8-nm-gate-long planar transistor demonstrates less than an (?) 1-mA/μm performance [15]. It is appreciably below the third limit. However, it can be exceeded by current boosting three-dimensional fin-gate structures.

Since the newest developed 3-D FinFETs have been announced reaching 1.0–1.5 mA/μm at V_{GS} of 0.75–0.8 V [16, 17], research work should concentrate on clarifying the risk of HCD in FinFET [18–20]. Inherent HCD risks in the FinFETs occur due to three reasons. First, the number of inversion electrons is increased by the surrounding three-dimensional gate overdriving ($V_{GS}-V_{TH}$), which supplies more electrons into the Si–H bond-breaking procedure through the multi-vibration mode. Second, the three-dimensional-surrounding gate introduces additional side interfaces between Si/SiO$_2$ by the fins, where additional Si–H bond breakage can occur. Furthermore, it can be enhanced when the gate is aligned to the (110) direction, the surface direction of the fin should also be (110) and the silicon surface density of the (110) plane is 1.4× larger than that of (100) plane [21]. Third, the three-dimensional surrounding of gate structures confines the heat dissipation only through the bottom-directional narrow body, thereby increasing the thermal resistance of the heat dissipation path. The lattice-carrier scattering generates heat that is referred to as "self-heating" and this increase of the lattice temperature is proportional to the thermal resistance. The temperature activation of HCD therefore becomes a critical reliability issue especially in high-current driving and poor heat dissipation devices like FinFETs [18]. A more detailed description of temperature dependency on hot carrier generations is found in the following section.

2.2. Temperature dependency on hot carrier generations

According to the LEM mechanism, carriers gain kinetic energy from the e-field, F, through free accelerated motion. The energy distribution of electrons is affected by the mean free path, λ, through,

$$f(E) = e^{-E/q\lambda F}. \tag{5}$$

Self-heating and/or ambient heating induces lattice vibrations that scatter the electrons to prevent gaining sufficient kinetic energy from triggering impact ionization. It can be assumed that λ decreases as the lattice temperature increases. As a result, higher energetic carriers can be generated at lower temperatures. **Figure 4** compares the long and high-biased (LH) transistors' and the short and low-biased (SL) transistors' HCI properties depending on temperature. That temperature dependency of the HCI lifetime and the substrate current follow the LEM picture in LH but not in SL suggesting that the LEM prediction is valid only in LH ranges but not in SL.

Monte Carlo simulation-based studies reveal that electron energy distribution function is composed of an e-field-driven main region and thermal tail [22, 23]. The knee voltage, $V_{EFF} = V_{DS} - V_{DSAT} + V_o$, separates two regions, where V_o is the voltage drop in the halo
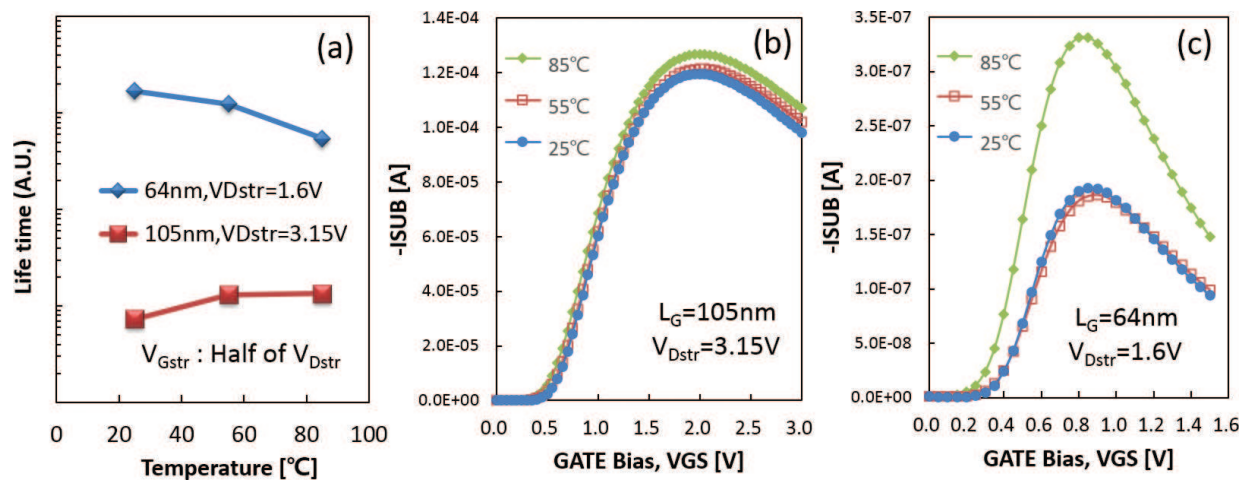
Figure 4. Hot carrier characteristics of 64-nm-gate-long, 1.6-V V_{DS} driving NMOSFET (SL) compares with that of 105-nm-gate-long, 3.15-V V_{DS} driving NMOSFET (LH). (a) Hot carrier lifetimes of SL and LH show different thermal dependencies: HC Lifetime of LH increases with temperature in accordance with the LEM framework, but does not for SL. (b) V_{GS}-dependent substrate currents of LH slightly enhance with temperature. (c) A remarkable increase at high temperature (85°C) found in SL suggests a temperature activation of HCD, which cannot be explained through the LEM framework.

region of the drain side. In LH transistors, hot carriers generated in the main region are dominant because of the large value of V_{DS}, which shows the negative dependency on temperature through λ. The scaled-down drain biases in short channel transistors reduce the V_{EFF} and the dominant hot carrier generation region is shifted from the main to the thermal tail via EES or MVHR. Since both are temperature-activating processes, the overall HCD shows a positive dependency on temperature that is detrimental especially in high-current-driving self-heating transistors like FinFETs.

2.3. PMOS hot carrier degradations

Traditionally, HCDs in PMOSFETs have not been taken seriously because a large energy barrier between Si and SiO_2 (~4.8 eV [24]) and a high-impact ionization threshold (Φ_{ii} = 1.43–1.92 eV [25]) of holes make difficult a LEM-like HCD in normal operational voltage ranges. The drain-avalanche hot-electron (DAHE) generates favorable electron injections into SiO_2 in low V_{GS} (1/3–1/4 of V_{DS}), which were known to be the dominant mechanism of HCD in PMOSFETs. The injected electrons fill the preexisting traps in the vicinity of the drain, which may cause effective gate length shortening, as $L_{eff} = L_o - \Delta L(N_{ox,e})$, and therefore punchthrough and breakdown may occur. However, this is a self-limiting procedure due to the exponential decrease of the electron injection current as a function of distance to the drain and hence yields a logarithmic dependency of $\Delta L(N_{ox,e})$ on time [26]. As PMOS gate oxide scales down, a turn-around of drain current degradation is observed, which is due to the charge re-emission and donor-like interface trap, $\Delta N_{it,d}$, generation under the high vertical field [25]. The dominant degradation driver has also been changed from the hot electron injection to the hot hole injection as Tox scales down. These transitions rely on (1) nitridation of gate oxide to suppress boron penetration, which enhances the generation of the positive charge (PC), (2) as oxide e-field, F_{ox}, exceeds 5MV/cm, NBTI degradation is triggered by cold hole injections at the source region, which are combined with hot hole injections at the drain region.

Abnormally large degradations of PMOS were reported in hot electron injection stress experiments at the cryogenic temperature of 77 K and subsequent anneal at elevated temperatures (300 K or higher) [27]. It is believed that the increase of carrier mobility and mean free path at 77 K creates additional damage sites in the oxide, which are initially inactivated at 77 K, and eventually convert to positive donor-type interface states as the de-trapped electrons leave vacancies in the annealing stage.

To summarize the hot carrier degradations in PMOSFETs, both electrons and holes created by impact ionization are responsible via their own natures for creating and/or changing the state of the oxide bulk traps and the interfacial traps. Since hot carrier generations in PMOSFETs are still negligible due to their low efficiency compared with those in NMOSFETs, cold hole injections to the SiO_2 can activate an appreciable number of interfacial and oxide bulk traps in the normal operational voltage range because holes are more efficient in trap generation processes than electrons are [28]. Cold hole injection is regarded as the most serious degradation mechanism of modern PMOSFETs. This subject is dealt with in Section 3 more precisely.

2.4. An energy-driven HCD modeling of NMOSFETs for circuit simulations

2.4.1. Aging model parameters

Transistors' degradation and the circuit performance degradation can be quantitatively analyzed through the circuit simulations by using the specific spice model parameter set, which we call "**aging parameters.**" Properly chosen aging parameters among the whole spice model parameters should be accurate over the full V_{DD} range varying V_{GS} and V_{DS} as a function of the "**age,**" which is an amount of "degradation." In summary, an age is accumulated during a prescribed operation time per transistor, the age shifts the aging parameters, and finally aging parameters reproduce the degraded transistors' characteristics. All the calculations are fulfilled during aging circuit simulations with self-consistent aging-parameter updates. A recursive process (age determines the degradation of transistor and vice versa) executes during the simulation. The complete sequence of the aging-parameters extraction and aging circuit simulations is schematically illustrated in **Figure 5**. Since the aging-parameters extraction is carried out under DC-stress conditions, some assumptions must be made regarding the validity of accumulated age and aging parameters updated during the AC circuit simulation, which include the following: (1) the static degradation rate and bias dependencies under DC stress conditions are assumed to be the same under AC stress conditions. This quasi-static approximation is generally accepted in HCD because the recovery after stress degradation is negligible and the total amount of degradation can be regarded as a singular function of AGE without any path dependencies; (2) the degradation is assumed to be a very slow process within the conventional time span of circuit simulation. It is an indispensable assumption for the sake of convenience and for the efficiency of the aging circuit simulation. It enables a decoupling of the aging accumulation from the aging-parameter update. One can accumulate age by using the voltage and current waveform, which is simulated with "constant" aging parameters during a prescribed time period, t_{CYC}. To control these non-overlapped sequences, aging circuit simulations can use two time variables, t and t_{AGE}. Age accumulation during t_{CYC} with time-invariant aging parameters is controlled by t. Aging parameters are subsequently updated by

Aging Parameters Extraction
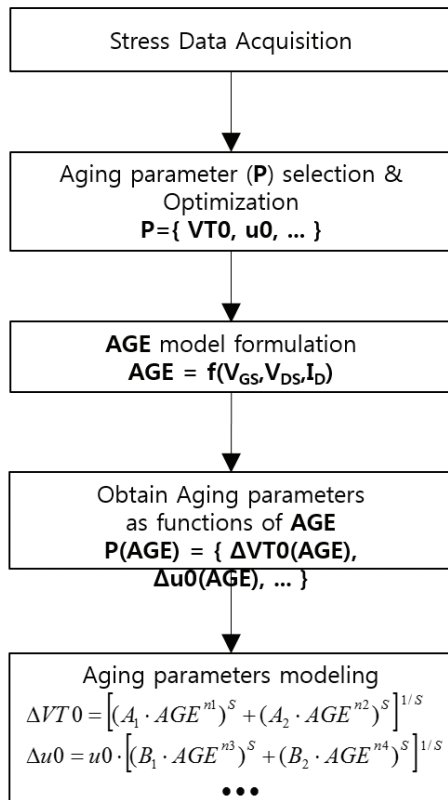
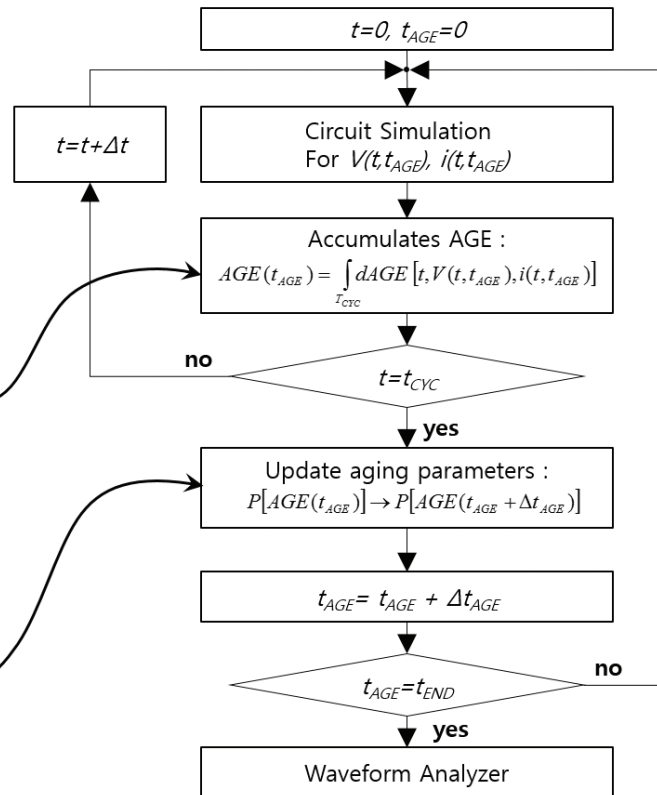## Aging Circuit Simulation

Figure 5. Schematics for aging-parameters extraction and aging circuit simulation procedures: Aging parameters (P) are selected from the transistor model parameters set to represent the experimental degradation with relevant physical mechanisms (*see left*). A unified degradation is modelled by an AGE and in turn aging parameters are extracted as a functionality of AGE. The aging circuit simulation is comprised of two non-overlapping loops, which are controlled by t and t_{AGE}, respectively (*see right*). The inner loop accumulates AGE with varying $V_{GS}(t)$ and $V_{DS}(t)$ for each transistor, and the outer loop updates the aging parameters using the accumulated AGE.

using the accumulated age as functions of t_{AGE}. A flowchart depicted in **Figure 5** (right) illustrates the sequence of the aging circuit simulation in detail.

The proper sequence for the aging-parameter extraction can be exemplified in the following example: a 62-nm-gate-long NMOSFET is DC stressed with a V_{DS} within the range of 2.1–2.3 V and a V_{GS} within the range of 1.5–2.3 V. After a 300-s stress, transistors $I_D − V_{GS}$, and $I_D − V_{DS}$ are typically compared to a fresh one as in **Figure 6**.

A threshold voltage shift and transconductance, Gm reduction, are found in stressed $I_D − V_{GS}$ and Gm $− V_{GS}$ as shown in **Figure 6(a)**, and **(b)**. Selecting the spice model parameters **VTH0** for threshold voltage shift and **u0** for Gm reduction is the obvious choice for the aging parameters since $Gm = \frac{W}{L} C_{ox} V_{DS} \frac{\partial \mu}{\partial V_{GS}}$ in low V_{DS} and only $\frac{\partial \mu}{\partial V_{GS}}$ term can be degraded by hot carriers. As a coefficient of the mobility model, **u0** can scale both μ and $\frac{\partial \mu}{\partial V_{GS}}$ as shown in Eq. (6). One can find another important feature of degradation in **Figure 6(b)**; the reduction of the Gm-declining rate with V_{GS} is distinct. It is related to the increase of interface trap charges. They do screen more e-fields from the gate, hence the influence of the gate is reduced and the
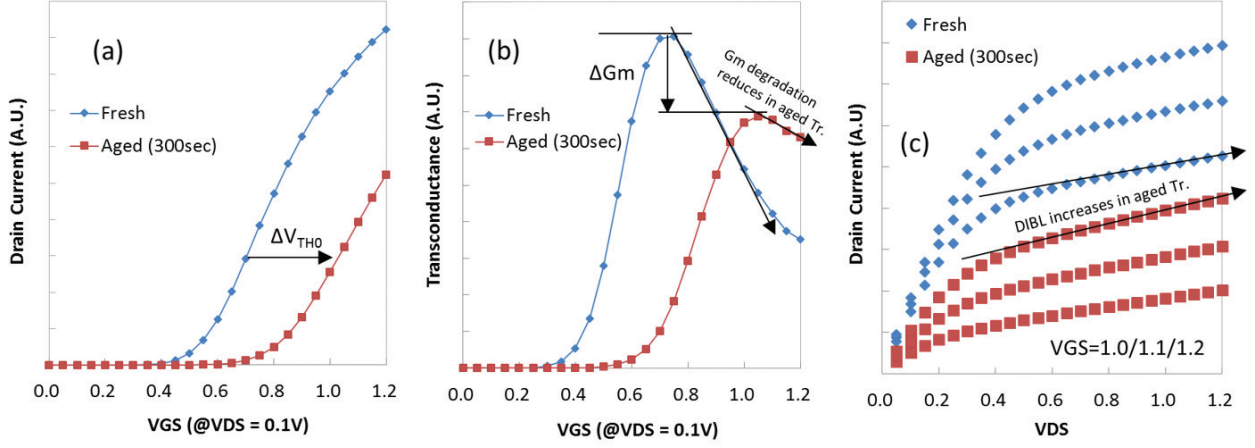
Figure 6. Comparisons of I_D-V_{GS}, Gm-V_{GS}, and I_D-V_{DS} showing the fresh and 300-s stressed with $V_{GS} = V_{DS} = 2.3$ V of 62-nm-gate-long transistors. (a) A comparison of fresh and aged transistors' I_D-V_{GS} clearly shows a stress induced **VTH0** shift. (b) Both the Gm and the Gm-declining rate on V_{GS} are reduced after being stressed. (c) A growing slope for the drain current on the V_{DS} is clearly shown in stressed I_D-V_{DS}'s, which can be interpreted as a DIBL increase due to HC stress.

surface-roughness-scattering-controlled Gm is less decreased in high V_{GS}. According to the spice model parameter equations, the Gm-declining rate can be modeled in the effective mobility, μ_{eff}, expressed in Ref. [29] as

$$\mu_{\text{eff}} = \frac{u0 \cdot f(L_{\text{eff}})}{1 + UA\left(\frac{V_{\text{gsteff}}+2V_{\text{th}}}{TOXE}\right) + UB\left(\frac{V_{\text{gsteff}}+2V_{\text{th}}}{TOXE}\right)^2 + UD\left(\frac{V_{\text{th}} \cdot TOXE}{V_{\text{gsteff}}+2\sqrt{V_{\text{th}}^2+0.0001}}\right)^2} . \qquad (6)$$

UA or UB may adjust the declining rate on V_{GS} (V_{gsteff} in Eq. (6)). But it is not preferable as both **u0** and UA (or UB) appear in the same model equation, which makes it difficult to extract their optimum values independently. In other words, a lack of orthogonality may affect the quality of the parameter extractions. Thus, an alternative choice can be **rdsw**, which is a spice model parameter expressing the extrinsic resistance of drain and source regions. The drain resistance is increased by the accumulation of trapped electrons in the drain region. The Gm-declining rate is also affected by the accumulation of trapped electrons, which screen the gate electric field. Thus, choosing **rdsw** can include both a drain resistance increase and a Gm-declining rate decrease of degraded transistors without any ambiguity among parameters. The last parameter can be determined by observing **Figure 5(c)**. As the drain current-increasing slope along the V_{DS} is clearly shown in the aged transistors, one can choose a DIBL (drain-induced barrier-lowering) control parameter. An increase of DIBL originates from the same mechanism, which causes the increase of **rdsw**; the vertical e-field is screened by trapped charges and hence the channel inversion charges become more susceptible to the lateral e-field or V_{DS}. The DIBL formulation in the spice modeling [29] is

$$\Delta V_{\text{th}}(DIBL) = -\theta_{\text{th}}(DIBL) \cdot (ETA0 + ETAB \cdot V_{BS}) \cdot V_{DS} . \qquad (7)$$

In Eq. (7), **ETA0** is a suitable parameter to describe the hot carrier-induced DIBL increase. Note that even though **VTH0** and **ETA0** may appear in the same threshold voltage model, they can

be distinguished from each other since **VTH0** is extracted from HC degradation data without any dependency on V_{DS}, but **ETA0** is the coefficient of V_{DS} in ΔV_{th} that implies that one can extract both **VTH0** and **ETA0** independently. Selected aging parameters, **VTH0**, **u0**, **rdsw**, and **ETA0,** are optimized via appropriate numerical processes to best fit the experimental data. **Figure 7** compares the results where points mark the experimental data and lines are spice simulation results using optimized aging parameters.

2.4.2. An energy-driven AGE model

The AGE is a commonly used parameter to accumulate the amount of degradation under various bias conditions in aging circuit simulation. Appropriate AGE model reflects underlying physics with a relevant functional form for the bias and time. According to field-driven HCD, one can define the AGE function as

$$AGE_{FD} = \frac{I_D}{W \cdot H} \left(\frac{I_{SUB}}{I_D} \right)^m \cdot t_{STR} \qquad (8)$$

where m is known to be around 3 and H is a constant according to the field-driven HCD framework. **Table 1** checks the validity of this assumption. In this table, we can find that the largest VTH0 degradation occurs when $V_{DS} = V_{GS}$, among the various V_{DS}/V_{GS} bias sets, while the maximum substrate currents and AGEs do not coincide with that of VTH0 degradation. This mismatch implies that the field-driven mechanism is no longer valid for the 62-nm-scaled NMOS transistor.

In order to adjust the discrepancy of the field-driven AGE, one can modify H to be a function of V_{DG}, as a commonly used relief in the field. The fitting results will be compared with newly developed energy-driven AGE's results later. A simplified version of the energy-driven AGE model is proposed in Ref. [30] as
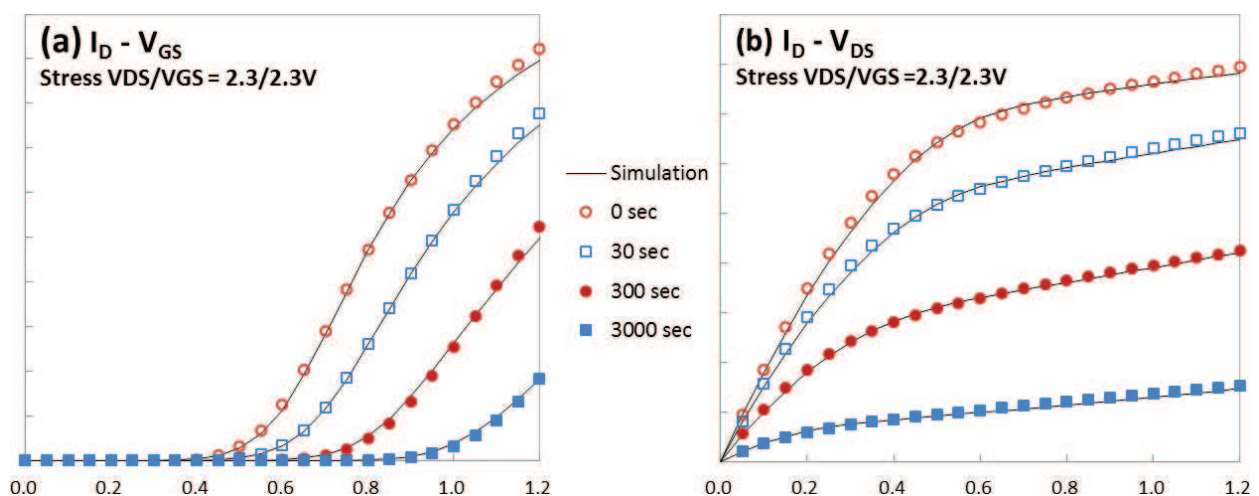

Figure 7. The aging-parameters optimization results fit the HCD measurements (points) with simulations (line) in (a) I_D-V_{GS} and (b) I_D-V_{DS} curves.

V_{DS}/V_{GS}	V_{DG}	I_{SUB}/I_D (x10^{-3})	AGE_FD ($H = 1$, $m = 3$)	Measured ΔVTH0 [mV]
2.1/2.1	0.0	0.491	3.32E-09	**151**
2.2/2.2		0.861	1.72E-08	**252**
2.3/2.3		1.572	9.97E-08	**395**
2.1/1.7	0.4	0.730	8.76E-09	76
2.2/1.8		1.210	3.97E-08	146
2.3/1.9		1.727	**1.01E-07**	257
2.1/1.3	0.8	**1.034**	**1.48E-08**	16
2.2/1.4		**1.754**	**8.44E-08**	95
2.3/1.5		**1.894**	8.61E-08	129

The maximum ΔVTH0 occurs at $V_{DS} = V_{GS}$ (V_{DG} = 0V) for each V_{DS}. While I_{SUB}/I_D and AGEs do not follow ΔVTH0 tendencies on bias.

Table 1. Field-driven ages (Eq. (8)) are calculated from the measured I_{SUB}, I_D, and **VTH0** degradation (Δ**VTH0**).

$$R_{age} = a(V_G - V_{th})^P \cdot e^{B(V_D - V_{DSAT})} + b(V_G)^C \qquad (9)$$

where R_{age} is the accumulation rate of the age, which is expressed by the multiplication of carrier density, $a(V_G - V_{th})^P$, carrier energy, $e^{B(V_D - V_{DSAT})}$ and $b(V_G)^C$ term for a high V_G dependency. Compared with Eq. (8), the linear carrier density dependency of I_D is generalized in Eq. (9) as having a power-law dependency with exponent P, which reflects the relevant mechanism of HC generation: one for field-driven, two for EES, and 20 for MVHR for a wide range of gate lengths of MOSFETs' and drain bias. The exponential term for carrier energy reflects the energy distribution of the electrons as a function of drain overdriving voltage, $V_D - V_{DSAT}$. The last term is negligible since it becomes significant only if the V_G is larger than 3.0, which is beyond the normal operation range in modern technology. The saturation voltage, V_{DSAT}, is originally defined by the drain current saturation point in MOSFETs' $I_D - V_{DS}$ relations. The velocity saturation of mobile carriers causes the drain current saturation of the scaled MOSFETs. At the same time, V_{DSAT} in the carrier energy distribution function defines the threshold energy to HCD. Although the notation V_{DSAT} is commonly used to denote the two different mechanisms, the values of V_{DSAT} for both mechanisms need not be the same. This is shown in Ref. [31] that the substrate current starts from a smaller V_{DSAT} but is still proportional to V_{DSAT}; thus the drain voltage dependency has the form of $V_D - \eta V_{DSAT}$, where η is a fitting constant defined within 0-1. From this observation, V_{DSAT}, as the threshold of HCD, should be extracted from both drain current and substrate current measurements, as a form of V_{DSAT}, or alternatively a new energy-driven AGE can be modified as

$$AGE_{ED} = (V_G - V_{th})^{n_G} \cdot \left(\frac{K}{\ln(I_D/I_{SUB})}\right)^{n_D} \cdot t_{STR} \qquad (10)$$

where the well-known $I_{SUB}/I_D = C(V_D - \eta V_{DSAT})e^{-V_o/(V_D - \eta V_{DSAT})}$ form is used to replace $V_D - \eta V_{DSAT}$ by its simplified expression: $V_D - \eta V_{DSAT} \approx K/ln(I_D/I_{SUB})$ and the exponential function for the energy dependency is replaced by a power-law function because it has a better

degree of freedom to fit to the experimental data. The last term of Eq. (9) is omitted as stated above. **Figure 8** compares the fitting results of the **ΔVTH0** by using the conventional e-field-driven AGE, Eq. (8), and the newly defined energy-driven AGE, Eq. (10). The H parameter in Eq. (8) is modified to have an exponential functional dependency of V_{GD} to fit the measurement data. As shown in the figure, the **ΔVTH0** tends to slow down as AGE increases. The saturation phenomenon is commonly found in the aging-parameter measurements. The interpretation for the saturation is that the current is pushed down by the interface electrons at the lightly doped drain region, which reduces the interface trap influence on the drain current reduction [32], or a saturation of preexisting charge trapping [33] results in a two-slope shape on the aging parameters dependent on the AGE. A behavioral expression of this effect can be generalized in the following expression:

$$\Delta P(\text{AGE}) = [(P_1 \cdot \text{AGE}^{n1})^S + (P_2 \cdot \text{AGE}^{n2})^S]^{1/S} \tag{11}$$

where S is the shape factor and has a negative digit. The two-slope combination of Eq. (11) is used to fit the **ΔVTH0** dependence on the AGE as shown in **Figure 8(b)**. As shown in the figure, the overall consistency is improved by using the energy-driven AGE defined by Eq. (10). **Figure 9** illustrates the aging-parameters extraction results by comparing the measurement and the aging simulation results. Two kinds of extraction methods are compared in the graph, which are as follows: (1) AGEs are extracted by using only fresh measurement values of V_{th}, I_D, and I_{SUB} and (2) AGEs are extracted by degraded V_{th}, I_D, and I_{SUB} in order to reflect "degradation of age" recursively. The overall matching property is improved by this update as shown in the figure.

2.5. A hot carrier-resistant design technique through V_{GS} ratio controls

2.5.1. V_{GS} ratio and ADF

The last example is to demonstrate a hot carrier-resistant design technique. HC-resistant design techniques have been attracting more attention as technology gets smaller. A strong demand can be found in typical DRAM word-line driver circuits where the inherent risk of
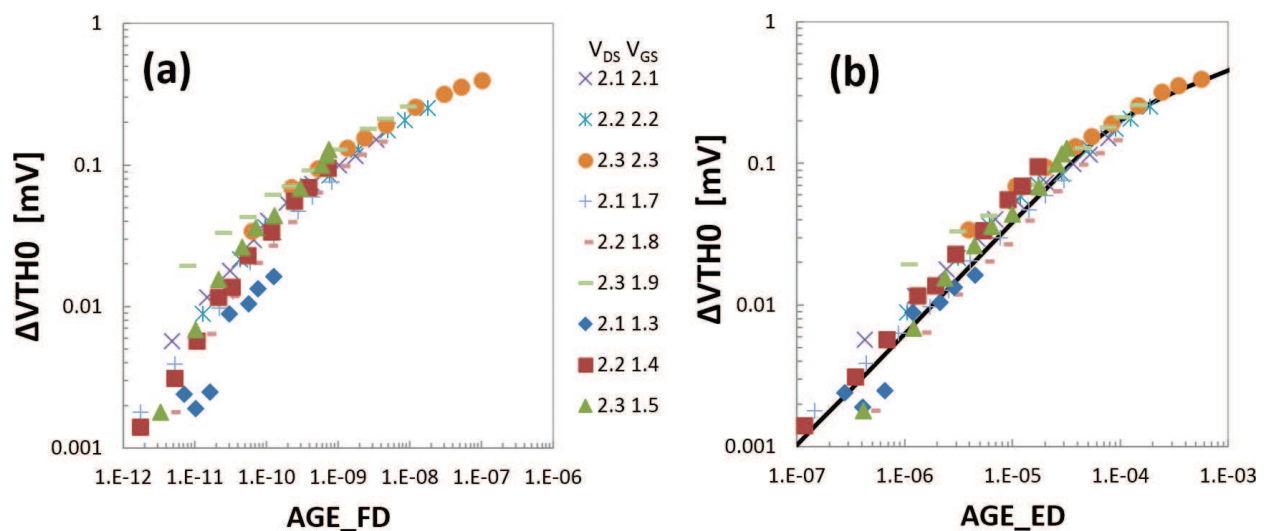

Figure 8. Optimized **ΔVTH0**s as a function of (a) the conventional field-driven age model (AGE_FD) and (b) the newly developed energy-driven age model (AGE_ED) are illustrated.
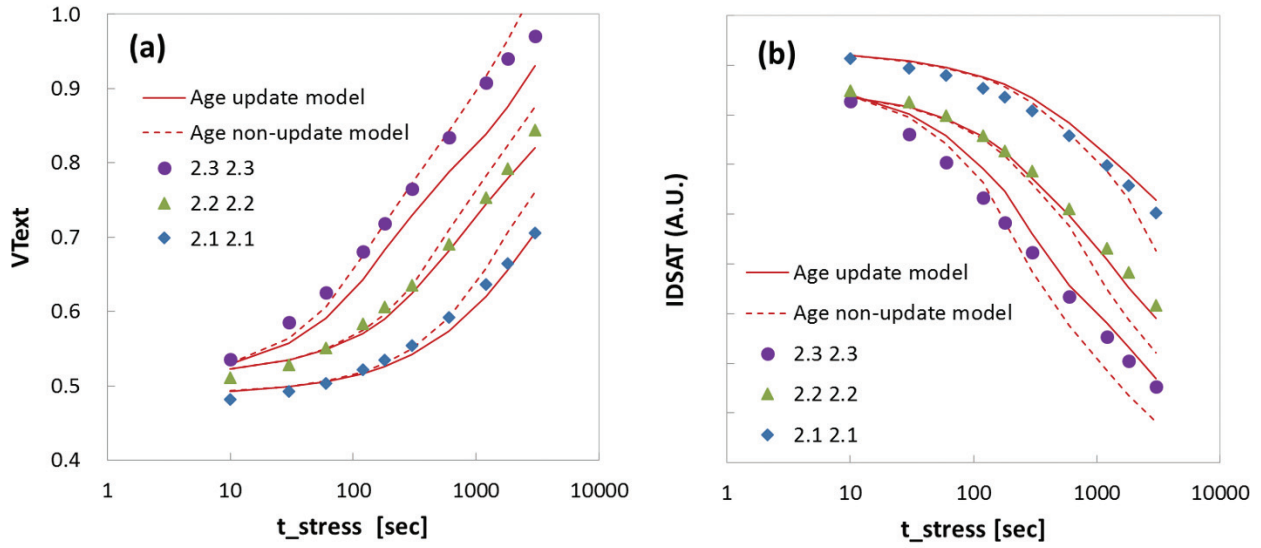
Figure 9. Aging-parameters' model extraction results are compared with the measurement data. As the age accumulation tends to slow down with the degrading ingredients (drain current and substrate current), the age update model parameter simulation results in (a) VText and (b) I_{DSAT} (solid lines) are less than those of the age non-update model (broken line). The age non-update model simulation results show an over-estimation especially in the deep degradation regime ($t_{STRESS} > 1000$ s). From this aspect, one can conclude that the overall accuracy of the simulation is improved by age update feedback.

HCD exists due to non-scale-down word-line-pumping voltage (V_{PP}). As stated above, the necessity for sustaining channel conductance in scaled cell transistors forces the V_{PP} to fix around 3 V. The field-driven mode can be a dominant HCD mechanism in such a high V_{PP}-biased 100-nm-long gate length transistors. According to the LEM, the maximum degradation occurs at the peak substrate current (I_{SUB}) generating V_{GS} condition. The peak I_{SUB} generation V_{GS} defines the "V_{GS} ratio", γ, which is $V_{GS,\,peak}/V_{DD}$. If the constant V_{DS} is applied under DC bias conditions, γ has its maximum value of around 0.5–0.6. The minimum value might be 1/3 since most CMOS transistors have their threshold voltage of 1/3 of V_{DD} and the drain voltage is assumed to pull down immediately as the gate turns on in the CMOS inverter operation. **Figure 10(a)** shows the HCD measurements with different V_{GS} ratios. The DC HCD reaches up to 40% at 100-h stress with $\gamma = 0.5$–0.6. Such a severe degradation does not seem to guarantee the lifelong serenity of the circuits without any kind of mitigation strategies. Several significant features of HCD are found in the figure:

1. As two time-slope phenomena are found in the figure as stated above, a sufficient timing margin is required to survive the initial rapid degradation. The overall timing shift of the WL driver reflects the "quasi-saturation effect" of the transistors' degradations as depicted in **Figure 10(b)**. As one can find in the figure, WL-off-degradation progresses toward a saturation at 5 years and very slowly degrades during further aging. In this design, wear-out is remarkably retarded due to this quasi-saturation phenomenon.

2. The V_{GS} ratio determines the quasi-saturation level of degradation. Compared at 100-h stress, only 1/3–1/2 of all degradations are shown in the $\gamma = 1/3$ stress condition as opposed to in $\gamma = 0.5$–0.6. Thus, the reduction of γ is the primary design target for long-term HCD reliability.
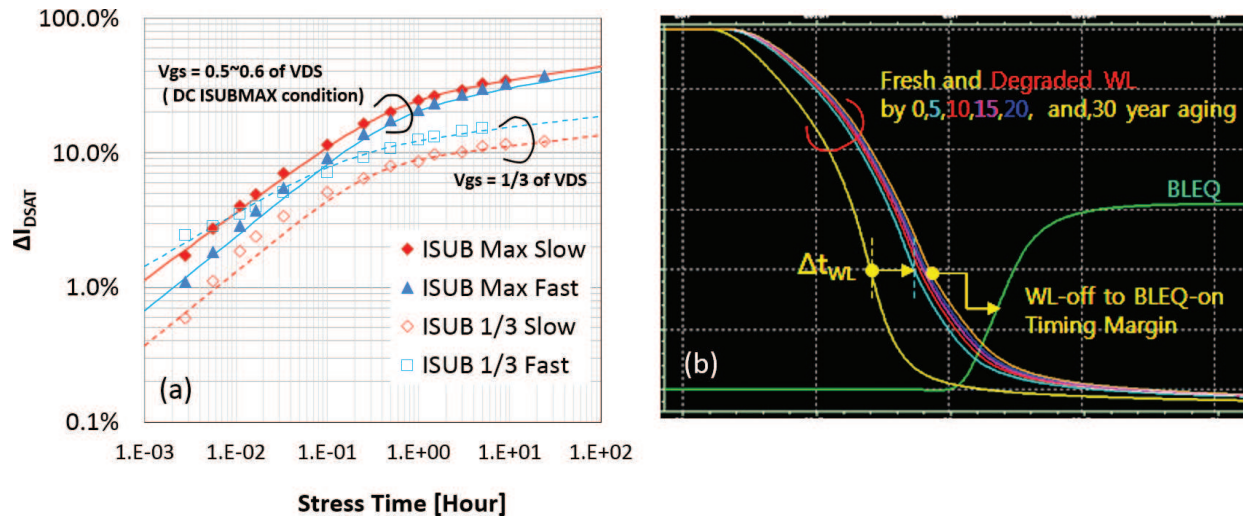
Figure 10. (a) I_{DSAT} degradations by HCI are depicted as functions of stress time. Two different choices of V_{GS} ratio affect amount of degradation and its quasi-saturation levels. The conventional choice of V_{GS} to let the substrate currents have their maximum among the V_{GS} values yields the maximum stress degradation of I_{DSAT} as shown in the figure. (b) The transistors' quasi-saturations influence on the word-line driver pull-down signal delays. The aging simulation by using the aging parameters reveals rapid degradations of word-line pull-down delay in the initial stress period (from 0 to 5 years) and their slowing down due to quasi-saturation in the subsequent period (from 5 to 30 years). It suggests that hot carrier-resistant design requires a sufficient timing margin to survive initial rapid degradation before the quasi-saturation takes place.

3. Process skews, which are mainly V_{TH} variations, affect HCD in a complicated manner. In the large γ case, a lower V_{TH} wafer (fast skew) has a smaller degradation than that of higher V_{TH} wafers (slow skew), and reverse for the small γ case. These phenomena can be explained by the two competing processes: (i) more gate overdriving, $V_{GS} - V_{TH}$ with lower V_{TH} reduces the lateral e-field and hence HCD, which predominates in the large γ case, (ii) lower V_{TH} increases the drain current as more electrons participate in the impact ionization process and hence cause higher HCD, which predominates in the small γ case. Thus, V_{TH} controls in order to improve HCD may cause the opposite results depending on γ.

As shown in **Figure 10**, we can conclude that the V_{GS} ratio has an effect not only on degradation rate but also on its quasi-saturation level by ×2–×3 differences. The quasi-saturation level of the degradation is especially important in long-term HCI degradation where most transistors suffer sufficient stress to enter the quasi-saturation region.

AC duty factor (ADF) is commonly used to estimate the HCD in AC operations. It is defined by the AGE in DC bias conditions divided by the aggregation of the AGE per circuit operation cycle. The commonly used form of ADF is shown as

$$\text{ADF} = \frac{I_{\text{SUB\_DC}}{}^m \cdot t_{\text{RC}}}{\int_0^{t_{\text{RC}}} I_{\text{SUB}}(t)^m \cdot dt} \quad . \tag{12}$$

As a large value for ADF improves AC hot carrier reliability, a straightforward HCI-resistant design may focus on as large a value as possible. However, this is not a necessary condition for long-term HCI-resistant design. **Figure 11** illustrates the HCD of critical transistors consisted
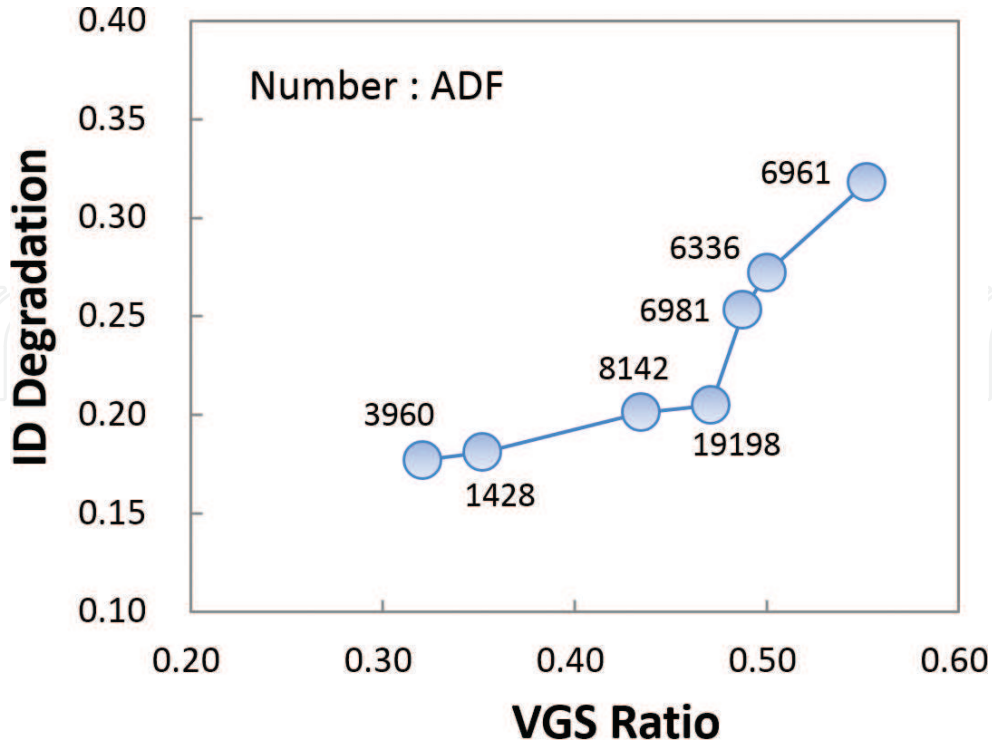
Figure 11. Transistor's current degradations are simulated to reproduce transistor AC degradation after a 10-year operation of a WL driver circuit. Simulation results are depicted with respect to the V_{GS} ratio. AC duty factors (ADFs) are also marked in the same graph as numbers.

of the word-line driver circuitry after a 10-year aging period, which is long enough to enter the quasi-saturation regime for all transistors. The significant role of the V_{GS} ratio can be reconfirmed in the figure. ADF may retard degradation while it has no influence on ΔI_D once it enters the saturation region. From this observation, we can conclude that reducing the V_{GS} ratio has a significant effect on the long-term hot carrier reliability. By contrast, a large ADF retards the HCD to reach its quasi-saturation level, but has no effect on reducing the quasi-saturation level itself.

Driver strength is the most important control parameter regarding circuit level HCI degradation. Strong drivers can easily pull down the output voltage, V_{DS}. Due to the exponential dependency on V_{DS}, the substrate current quickly diminishes with the fast pull-down and, as a consequence, ADF increases and γ decreases, respectively. **Figure 12** illustrates a typical example of substrate current shape as a function of V_{GS} in a CMOS inverter circuit. By increasing driver strength, the substrate current peak is decreased and hence HCI stress is mitigated.

2.5.2. HCI-resistant design strategy

As shown above, reducing the V_{GS} ratio and increasing ADF can be recommended as HCI-resistant design strategies. Reducing input slew rate and increasing output slew rate or increasing driver size is a straightforward method to obtain both design targets. Reducing

V_{DS} humps through increasing gate-drain overlap capacitance of the strong driver is an additional benefit to mitigate HCD. One can find an example of HCI-resistant design for a typical inverter chain logic depicted in the inset of **Figure 13(a)**. As driver size splits from ×1 to ×2, ×4, and ×8, input and output slew rates are correctly modified and the I_{SUB} waveforms go
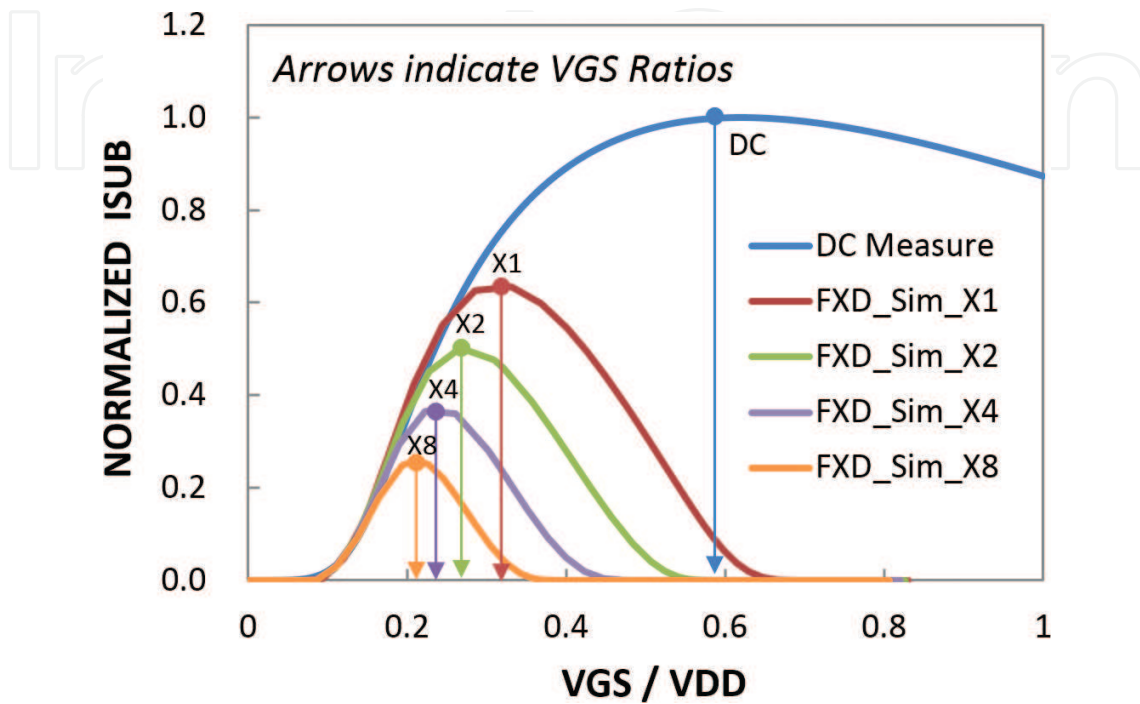

Figure 12. Substrate current profiles are compared as functions of normalized gate bias (V_{GS}/V_{DD}) in various driver strengths. The V_{GS} ratio is maximized in DC sweep mode (constant V_{DS} is applied) and decreased in CMOS inverter operation mode (increasing V_{GS} pulls down V_{DS}). The driver strength strongly influences both substrate currents and the V_{GS} ratios. From this context, one can deduce that the long-term hot carrier reliability can be improved with driver strength.
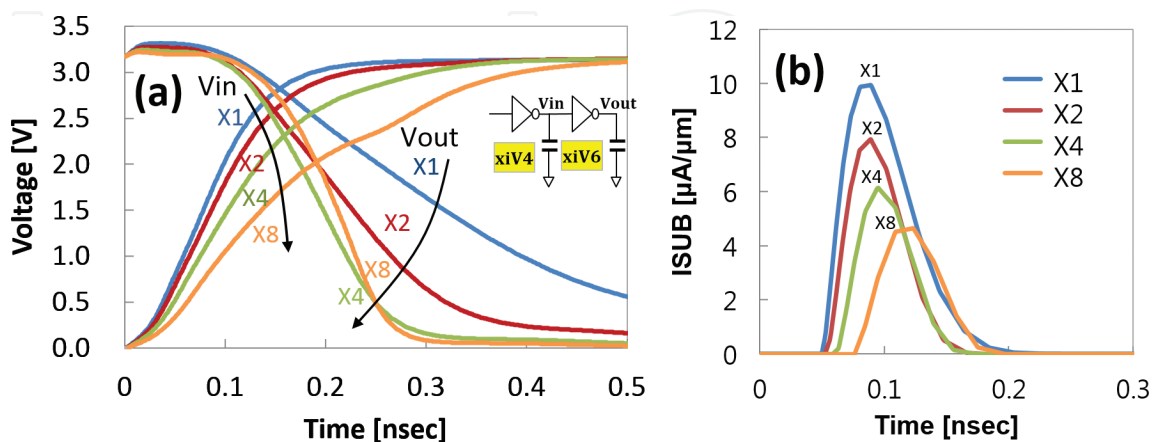

Figure 13. (a) Simulated input and output voltage waveforms of an inverter driver (xiV6), which are measured at the gate node and the drain node of xiV6, respectively. Input slews decrease and output slews increase with driver strength results in deductions in I_{SUB} as depicted in (b).

down as shown in **Figure 13(b)**. The HCD simulation for the inverter chain consisting of pre-driver (xiV4) and main-driver (xiV6) results is depicted in **Figure 14** with main driver size splits. The increase in size reduces HCD of the main driver as shown in the figure with a slight loss of that of the pre-driver. The increase in driver size, or fan-out, means more layout area consumption and a deterioration of pre-driver's HCD, and hence should be compromised. A trade-off design choice between HCI robustness and area penalty may exist within the range of ×2–×4 in the examples.

2.6. Summary

The hot carrier degradation mechanism has evolved from the single-particle e-field-driven model to the multi-particle energy or current-driven model associated with consistent technology scale-downs. During the last 30 years, a general agreement has been made that the complete extinguishing of this kind of catastrophic failure is impossible, though a better understanding of the physical mechanism and the relevant modeling contributes to developing HCI-aware design techniques. As a part of which the aging-parameters optimization of HCD of a 62-nm gate-long NMOS transistor has been demonstrated. Newly developed energy-driven AGE formulations show a better consistency with the experimental data than conventional e-field-driven AGE models without any arbitrary correction function of H, which implies that an energy-driven HCD predominates in the deca-nanometer-scaled transistors. A hot carrier-resistant design example for the DRAM word-line driver is also presented. From this study, a long-term hot carrier-resistant design strategy can be summarized as follows: (1) Give sufficient timing margins to survive the rapid initial degradation. (2) The V_{GS} ratio is the key control parameter since it directly relates to the quasi-saturation level of long-term HCD. (3) A prime design target is driver strength, or fan-out. Stronger drivers reduce the HCD through the increase of ADF and, more importantly, through the decrease of the V_{GS} ratio but at the cost of an area penalty. A compromise between driver strength and area penalty is a required trade-off in HCI-resistant design solutions.
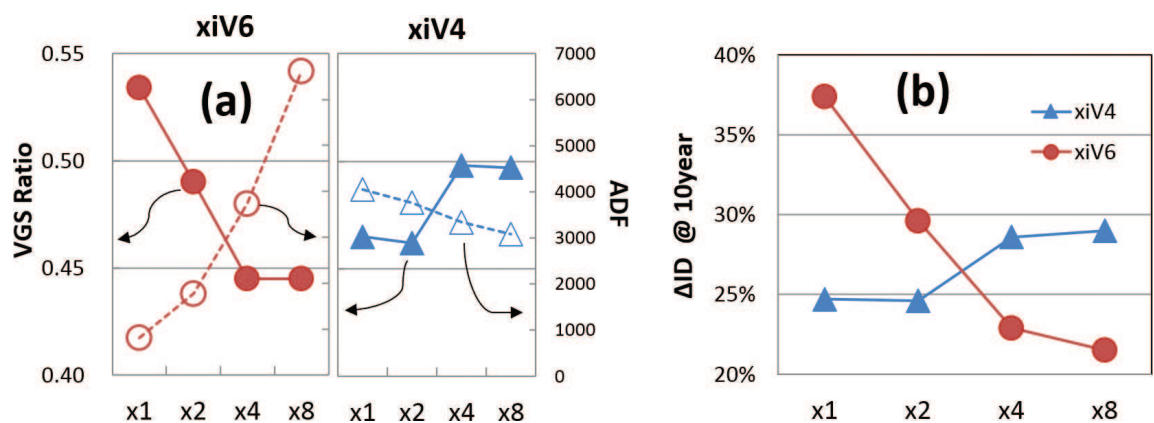

Figure 14. (a) Fan-out increasing enhances V_{GS} ratio and ADF in the xiV6 driver while both HCI-related parameters (V_{GS} ratio and ADF) deteriorate in the XiV4 pre-driver. (b) I_D degradation after a 10-year operation increases for xiV6, as opposed to decreases for xiV4. An optimal trade-off of driver strength (fan-out) considering both long-term HCI resistance and area penalty can be made in x2–x4 range of driver strength (fan-out).

3. Negative bias temperature instabilities (NBTI) in PMOSFETs

3.1. Introduction

As deca-nanometer-scaled transistors require the oxide scaled down to around 20 Å for sufficient gate control, a 5–10 MeV/cm of vertical e-field is easily established in the nitride-cooperated silicon dioxide region under the normal V_{DD} conditions. The Fowler-Nordheim tunneling mechanism can be triggered in such a high e-field, negative bias applied in PMOS gates collect inversion holes and then tunneling into gate oxide by e-field driving. Although this range of e-field is not sufficiently strong to generate hot carriers, cold holes in PMOS can interact with the oxide bulk traps and hydrogen passivated in the Si/SiO$_2$ interface with temperature activation, which in turn result in positive charges in the oxide region and the Si/SiO$_2$ interface. The critical characteristics of PMOS, threshold voltage (V_T), drain current (I_D), and transconductance (Gm), can be degraded by the trapped positive charges once the gate is negatively biased in moderate thermal conditions, regardless of V_{DS} or drain current. Owing to its nature, negative bias temperature instability (NBTI) has been an urgent issue in state-of-the-art PMOS transistors, which are prone to gate-tunneling hole-induced degradation. NBTI-resistant design is quite difficult because a simple turn-on operation triggers NBTI degradation. This means that the degradation occurs during the whole period of PMOS turn-on, and thus the only possible way to prevent it seems to be by a "power cut" during PMOS standby periods, which incurs spatial and performance penalties.

When the negative bias is applied, oxide bulk charges, which are called E' centers (oxygen vacancies) interact with the holes or protons (H$^+$) to produce positive charge build-up in the oxide bulk [34]. Interface trap (N_{it}) generation is attributed to a breaking of Si–H bonds by holes [35], or by protons [36], which leave behind the amphoteric trivalent $\text{Si}_3 \equiv \text{Si}^{\bullet}$ defects, that is, silicon-dangling bonds, which are called P_b centers. Thermal nitride deposition during gate oxidation has been known to cause additional profiles of the positive trap states, which are called DLHT [9, 10] or positive charges (PCs) [37]. When the stress bias turns off, a part of the PCs can be neutralized by the bulk electron from the N-Well, which causes a quick recovery after stress. The NBTI-degrading species and related mechanisms can be summarized as follows:

1. Gate oxide-injected cold holes create positive charges through dissociation of Si–H bonds at the interface (ΔN_{it}) or by being captured in the E' center in the oxide bulk, which is responsible for ΔN_{ot}.

2. Atomic hydrogen, H^+, or protons predominately supplied by end-of-line anneal steps can generate interface traps, N_{it} through the following chemical reaction [36]:

$$(\text{Si} - \text{H}) + \text{H}^+ \leftrightarrow (\text{Si}^*)^+ + \text{H}_2 \tag{13}$$

where (Si*) denotes the unbonded silicon lattice (P_b center) and the remaining symbols have their conventional meanings.

3. Molecular hydrogen is responsible for the annealing of E' traps through repassivation of one hydrogen atom from H$_2$, and the other diffuses away [34].

4. Positive charges (PCs) are generated during negative stress. Distinct increases of PCs are found in thermally nitrided oxide (TNO), which possibly originate from the emission of electrons from nitrogen donors [9]. A part of the PCs is quickly neutralized in the subsequent recovery cycle by electron injection from the N-Well.

Except for point (4), the interface trap charge generation, ΔN_{it} tends to be proportional to oxide bulk-charge generation, ΔN_{ot} [38, 39] since they commonly originate from the same species, hole and H^+. One may focus only on ΔN_{it} to understand the role of hydrogen upon NBTI. The positive charge generation process elucidated in point (4) can be neglected since it can be regarded as only a short-term process, which has little effect on the long-term reaction and diffusion of hydrogen. To show a practical case study in the subsequent section, one can draw a qualified image of NBTI through a reaction and diffusion model of hydrogen for a better understanding of NBTI-aware process integrations.

3.2. Reaction-diffusion framework

The generalized form of N_{it} rate equation is

$$\frac{dN_{it}}{dt} = k_F(N_0 - \Delta N_{it} - N_{it,0}) - k_R(\Delta N_{it} + N_{it,0}) \cdot H_0^{1/\alpha} \qquad (14)$$

where H_0 is the Si/SiO$_2$ interface concentration of hydrogen species (α = 1 for atoms, 2 for molecules), and N_0 is the Si/SiO$_2$ interface concentration of Si–H bonds, $N_{it,0}$ is the preexisting interface trap density and ΔN_{it} is the newly generated interface trap density, respectively. The dominant species of diffusion into the oxide and gate region is known to be molecular hydrogen, which can be governed by the following two-dimensional diffusion equation:

$$\frac{\partial H}{\partial t} = D_H \frac{\partial^2 H}{\partial x^2} + D_H \frac{\partial^2 H}{\partial y^2}. \qquad (15)$$

Since the N_{it} originates from the dissociation of hydrogen at the interface, the total amount of N_{it} is assumed to be equal to (1) the sum of the hydrogen species in the gate regions plus (2) the amount of hydrogen species diffused out from the gate regions. The numerical expression may be expressed as follows

$$N_{it} = \int_{\text{Gate Regions}} H(x, y, t) \cdot dxdy \quad + \int_{t'=0}^{t} \oint_{\text{bound}} k_p H(x, y, t') \cdot dS \cdot dt' , \qquad (16)$$

with the assumption that all the hydrogen species, which reach the boundary are absorbed at a surface absorption velocity, k_p [cm/s] is expressed as

$$-D_H \cdot \nabla H|_{\text{boundary}} = k_P H|_{\text{boundary}} . \qquad (17)$$

The existence of the ideal sink at the boundary is a rather unphysical assumption but its exclusion is unavoidable because it is impossible to remove the complicated ambient effect of

the outside hydrogen. Furthermore, it improves the feasibility of the simulation. The diffusion constant of the hydrogen species, D_H, has strong material dependencies as shown in Ref. [7], which indicates that the diffusion speed of neutral species (H0, H2) is highest in the oxide, next in the poly, next in Si substrate, and extremely slow in the nitride film. Since hydrogen diffusivities have strong dependencies on material and are assumed to also have strong dependencies on film deposition conditions and even on the structure, one may rely on a numerical optimization method, as the only feasible way to determine each parameter (D_H, k_F, k_R, k_p, and N_0). Although physical uncertainty may be caused by a numerical optimization and it can be a major drawback for the reaction-diffusion framework, it is still valid for the relative analysis based on comparisons between the experimental data with consistent usage of optimized parameters. Simulation results are compared to the experimental value in **Figure 15**. A timing exponent measured as 0.17 for 10–1000 s as shown in the figure suggests that the diffusion species is molecular hydrogen [40]. An increase of N_{it} rate for 1000–10,000 s is ascribed to the out-diffusion of hydrogen through boundary layers. The N_{it} rate goes down and finally saturates after 100,000 s as N_{it} approaches N_0, which was also predicted in [40].

3.3. End-of-line anneal effects on NBTI

The preexisting hydrogen, $H(x, y, 0) = H_{PRE}$ seems not to have any influence on the N_{it} slope because the diffusion equation of hydrogen (Eq. (15)) is independent of N_{PRE}. However, atomic hydrogen also generates N_{it} as well as holes do as indicated in Eq. (13). One can postulate that
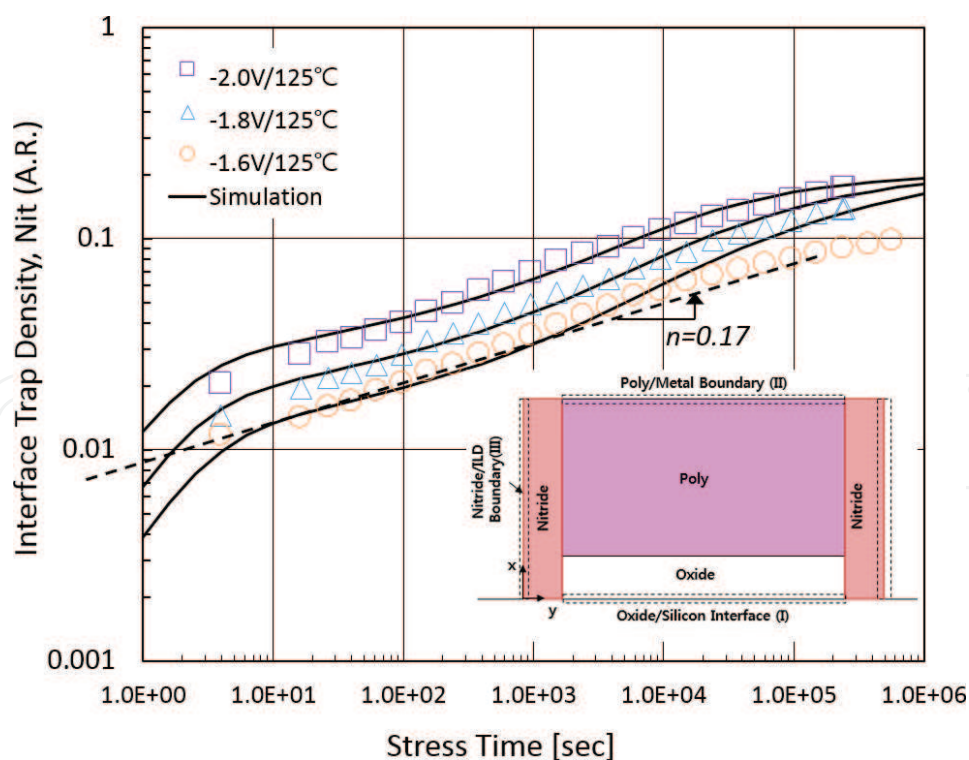

Figure 15. The simulation results of the reaction-diffusion of hydrogen are compared to measured NBTI degradations expressed as the interface trap generation rates. Inset: The simulation region consists of the oxide, poly, sidewall nitrides, and boundaries.

N_{PRE} can "increase" N_0 assuming that non-overlapped energy bands exist in the Si-H dissociation process caused by holes and by atomic hydrogen, respectively. Evidence for this assumption can be found in charge-pumping measurements and reproduced simulation results as illustrated in **Figure 16**. Interface trap densities, N_{it}, are measured by the charge-pumping method before and after 300-s-NBTI stress. The increase of N_{it} or ΔN_{it} after stress shows inversely proportional relations to the initial N_{it} or $N_{\text{it},0}$ as shown in the figure. One can interpret this dependency through Eq. (14) as the forward reaction term decreases as an increase of $N_{\text{it},0}$ and the reverse reaction (repassivation of the interface trap) term increases with $N_{\text{it},0}$.

An intriguing aspect is found in parameter fitting for sample A, B1, and B2: using a higher value of N_0 is indispensable for sample A to fit the experimental ΔN_{it} rather than for samples B1 and B2. This difference is assumed to originate from different EOL anneal condition splits for samples A, B1, and B2. All the samples undergo the hydrogen passivation process in a very low-pressurized anneal chamber with N_2 ambient at temperatures of 365–390°C (B2 sample experiences higher temperatures than others). Additional forming gas ($H_2 + N_2$) anneal applies only in sample A in the atmospheric chamber with 390°C. The precedent anneal applied to all the samples can passivate the silicon-dangling bonds with hydrogen, which is believed to come from the hydrogen-rich passivation layer that covered all the wafers, and the remaining hydrogen species may diffuse out from the silicon due to the
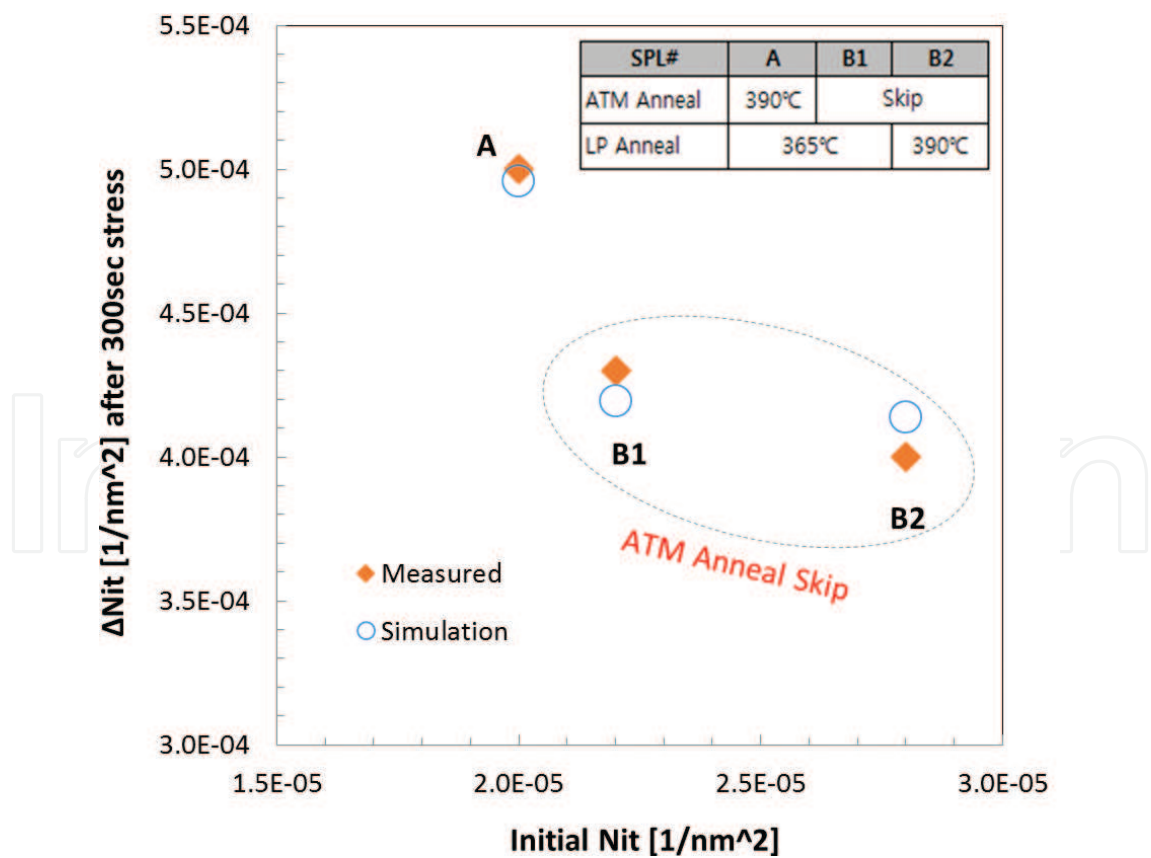

Figure 16. Charge-pumping measurement data before and after 300-s-NBTI-stress are compared to the R-D simulation results.

low-pressure ambience. Additional EOL anneal applies only to sample A, which can passivate additional silicon-dangling bonds and may leave a volume of hydrogen species in the gate regions. It causes additional passivation to reduce initial interface trap density, $N_{\mathrm{it},0}$ and to increase ΔN_{it} after NBTI stress by the remaining hydrogen, which can be expressed as the increase of N_0 in Eq. (14).

The evidence of remaining hydrogen, which is supposedly interstitial hydrogen, and induced N_0 enhancement, is also found in the anneal time split results shown in **Figure 17**. The low-pressure anneal time is doubled in the split group and it shows an earlier saturation of ΔN_{it} than the control group as shown in **Figure 17**. More interstitial hydrogen can diffuse out during the extended anneal step. It is believed that the reduction of interstitial hydrogen through the extended low-pressure anneal can reduce ΔN_{it} during NBTI stress. This is also reproduced in the simulation with two assumptions: a 25% increase in N_0 and the preexistence of hydrogen for the control group as compared in **Figure 17**. This conflict results in passivation anneal splits; ΔN_{it} increases with the additional atmospheric anneal, but decreases with the additional low-pressure anneal, which strongly suggests that the remaining hydrogen can make additional Si–H bond breakage plus that which the holes do. This can be reproduced by reaction-diffusion simulations through the increase of N_0 and H_{PRE}. From this plausible interpretation, one can conclude that removing hydrogen as much as possible from the transistor gate regions improves the long-term NBTI reliability.
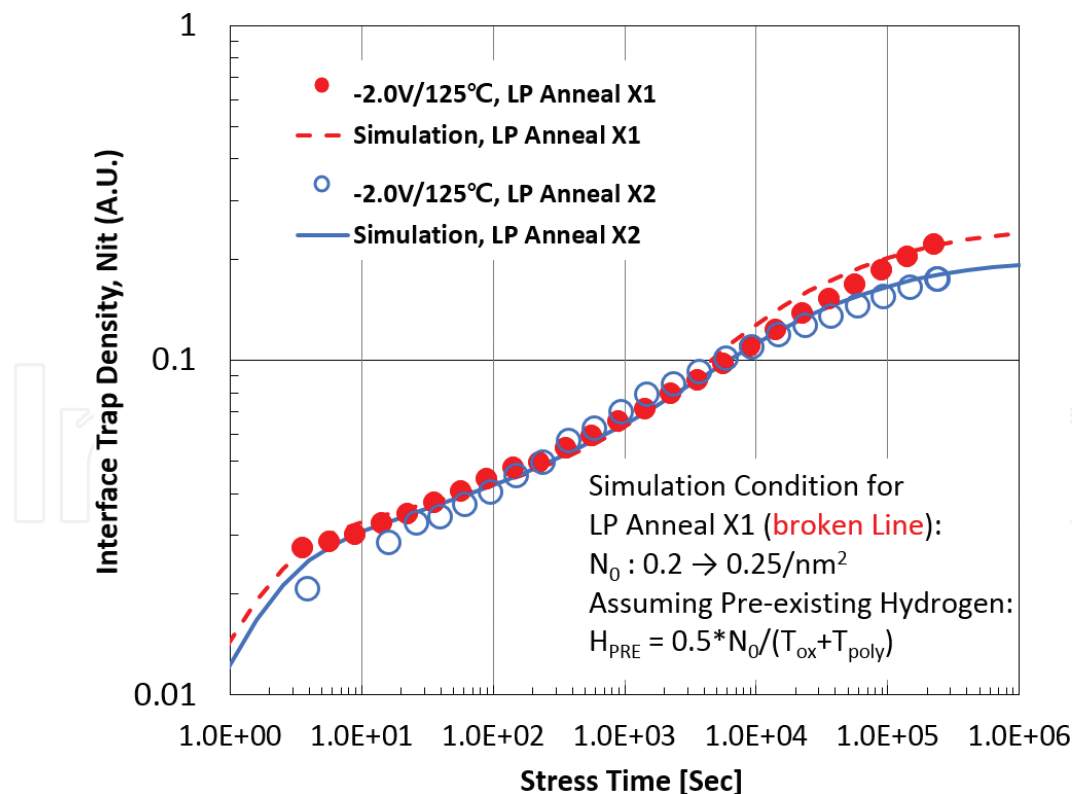

Figure 17. Interface trap generation rates of the low-pressure anneal split samples are compared to the R-D simulation results. Incomplete anneal split (LP Anneal X1 depicted in red line) is modelled by (1) increasing N_0 from 2.0 to 2.5 /nm² and (2) assuming pre-existing hydrogen in the gate regions.

3.4. Summary

NBTI has become the predominate long-term reliability threat as gate oxide is scaled down to the 20-Å range. It is caused by various source species: channel hole injections into gate oxide breaks the Si–H bond. Preexisting atomic hydrogen also dissociates Si–H bond to form a molecular hydrogen. Oxide bulk traps or nitride traps can be activated by holes or atomic hydrogen captures, or emitting electrons. Molecular hydrogen can neutralize the oxide bulk traps. Positive charges quickly generate and disappear at the initial stage of stress and recovery, respectively, which is believed to be associated with the nitrogen donor traps. The reaction-diffusion model simulation has demonstrated to reproduce the experimental ΔN_{it} data with the following features: a power-law functional dependency of 0.17 timing exponent for 10–1000 s and an eventual increase and saturation during the subsequent stress period. It proves that the reaction-diffusion framework of hydrogen, although still controversial among researchers, evidently reproduces NBTI degradation characteristics. Another R-D analysis of hydrogen is demonstrated in the EOL anneal experiments, which reveal that (1) preexisting hydrogen cooperates with holes to break the Si–H bonds, which can be modeled by an increase of the total number of breakable Si–H bonds, N_0. (2) Removing unbonded hydrogen from the transistor gate improves the long-term NBTI reliability.

Acknowledgements

Author details

Chang Yeol Lee

Address all correspondence to: changyeol.lee@sk.com

SK hynix, Icheon, South Korea

References

[1] Intel Chips timeline. [Internet].2012. Available from: http://www.intel.com/content/www/us/en/history/history-intel-chips-timeline-poster.html [Accessed: 2017-05-08]

[2]   Bornstein W, et al. Field degradation of memory components from hot carriers. In: IEEE 44th Annual International Reliability Physics Symposium (IRPS); San Jose; 2006. pp. 295-305

[3]   Hu C, et al. Hot-electron-induced MOSFET degradation—Model, monitor, and improvement. IEEE Journal of Solid-Sate Circuits. 1985;**SC-20**(1):295-305

[4]   El-Mansy YA, Boothroyd AR. A simple two-dimensional model for IGFET operation in the saturation region. IEEE Transactions on Electronic Devices. 1977;**ED-24**(3):254-262

[5]   Hess K, et al. Giant isotope effect in hot electron degradation of metal oxide silicon devices. IEEE Transactions on Electronic Devices. 1998;**45**(2):406-416

[6]   Li E, et al. Projecting lifetime of deep submicron MOSFETs. IEEE Transactions on Electronic Devices. 2001;**48**(4):671-678

[7]   Krishnan AT, et al. Material dependence of hydrogen diffusion: Implications for NBTI degradation. In: International Electron Devices Meeting (IEDM); Washington, DC, USA, 2005. pp. 4-7

[8]   Kufluoglu H, et al. A geometrical unification of the theories of NBTI and HCI time-exponents and its implications for ultra-scaled planar and surround-gate MOSFETs. In: International Electron Devices Meeting (IEDM); San Francisco, CA, USA, 2004. pp. 113-116

[9]   Ang DS, et al. Consistent deep-level hole trapping model for negative bias temperature instability. IEEE Transactions on Device and Materials Reliability. 2008;**8**(1):22-34

[10] Ang DS, et al. Effect of hole-trap distribution on the power-law time exponent of NBTI. IEEE Electronic Device Letters. 2009;**30**(7):751-753

[11] Randriamihaja YM, et al. New hot carrier degradation modeling reconsidering the role of EES in ultra short N-channel MOSFETs. In: IEEE International Reliability Physics Symposium (IRPS); Monterey, CA, USA, 2013. pp. XT.1.1-XT.1.5

[12] Rauch SE, III, et al. Role of E-E scattering in the enhancement of channel hot carrier degradation of deep-submicron NMOSFETs at high VGS conditions. IEEE Transactions on Device and Materials Reliability. 2001;**1**(2):113-119

[13] Bravaix A, et al. Hot-carrier to cold-carrier device lifetime modeling with temperature for low power 40 nm Si-bulk NMOS and PMOS FETs. In: International Electron Devices Meeting (IEDM); Washington, DC, USA, 2011. pp. 622-625

[14] Guerin C, et al. The energy-driven hot-carrier degradation modes of nMOSFETs. IEEE Transactions on Device and Materials Reliability. 2007;**7**(2):225-235

[15] Suk SD, et al. Characteristics of sub 5nm tri-gate nanowire MOSFETs with single and poly Si channels in SOI structure. In: Symposium on VLSI Technology (SOVT); Kyoto, Japan, 2009. pp. 142-143

[16] Cho H-J, et al. Si FinFET based 10 nm technology with multi V_t gate stack for low power and high performance applications. In: Symposium on VLSI Technology (SOVT); Honolulu, HI, USA, 2016

[17] Guo D, et al. FINFET technology featuring high mobility SiGe channel for 10nm and beyond. In: Symposium on VLSI Technology (SOVT); Honolulu, HI, USA, 2016

[18] Khan MI, et al. Self-heating and reliability issues in FinFET and 3D ICs. International Conference on Solid-State and Integrated Circuit Technology (ICSICT); Guilin, China, 2014

[19] Miyashita T, et al. High voltage I/O FinFET device optimization for 16nm system-on-a-chip (SoC) technology. In: Symposium on VLSI Technology (SOVT); Kyoto, Japan, 2015. pp. T152-T153

[20] Jin M, et al. Reliability characterization of 10 nm FinFET technology with multi-VT gate stack for low power and high performance. In: International Electron Devices Meeting (IEDM); San Francisco, CA, USA. 2016. pp. 380-383

[21] National Chiao Tung University, Taiwan. Information Technology Service Center. National Chiao Tung University. Silicon Surface Density [Internet]. 2017. Available from: http://web.it.nctu.edu.tw/~thhou/09-dee4515/Silicon%20Surface%20Density.pdf [Accessed: 2017-05-08]

[22] Bude JD, et al. Impact ionization and distribution functions in sub-micron nMOSFET technologies. IEEE Electronic Device Letters. 1995;**16**(10):439-441

[23] Zanchetta S, et al. Analytical and numerical study of the impact of HALOs on short channel and hot carrier effects in scaled MOSFETs. Solid-State Electronics. 2002;**46**:429-434

[24] Hofmann KR, et al. Hot-electron and hole-emission effects in short n-channel MOSFET's. IEEE Transactions on Electronic Devices. 1985;**ED-32**(3):691-699

[25] Bravaix A, et al. Hole injection enhanced hot-carrier degradation in PMOSFETs used for systems on chip applications with 6.5-2 nm thick gate-oxides. Microelectronics Reliability. 2004;**44**:65-77

[26] Brox M, et al. A model for the time- and bias-dependence of p-MOSFET degradation. IEEE Transactions on Electronic Devices. 1994;**41**(7):1184-1196

[27] Song M, et al. Comparison of NMOS and PMOS hot carrier effects from 300 to 77K. IEEE Transactions on Electronic Devices. 1997;**44**(2):268-276

[28] Brox M, et al. Dynamic degradation in MOSFET's—Part I: The physical effects. IEEE Transactions on Electronic Devices. 1991;**38**(8):1852-1858

[29] Dunga MV, et al. BSIM4.6.0 MOSFET model – user's manual. Department of Electrical Engineering and Computer Sciences. University of California, Berkeley; 2006

[30] Arfaoui W, et al. Application of compact HCI model to prediction of Process effect in 28FDSOI technology. In: IEEE International Integrated Reliability Workshop (IIRW); S. Lake Tahoe, CA, USA, 2014. pp. 69-72

[31] Yang JJ, et al. A new approach to modeling the substrate current of pre-stressed and post-stressed MOSFET's. IEEE Transactions on Electronic Devices. 1995;**42**(6):1113-1119

[32] Dreesen R, et al. Modelling hot-carrier degradation of LDD NMOSFETs by using a high-resolution measurement technique. Microelectronics Reliability. 1999;**39**:785-790

[33] Lachenal D, et al. HCI degradation model based on the diffusion equation including the MVHR model. Microelectronic Engineering. 2007;**84**:1921-1924

[34] Tahanout C, et al. Oxide trap annealing by H2 cracking at E' center under NBTI stress. In: International Conference on Microelectronics (ICM); Algiers, Algeria, 2012

[35] Fujieda S, et al. Interface defects responsible for negative-bias temperature instability in plasma nitride SiON/Si (100) systems. Applied Physics Letters. 2003;**82**(21):3677-3679

[36] Tsetseris L, et al. Physical mechanisms of negative-bias temperature instability. Applied Physics Letters. 2005;**86**:142103-142105

[37] Hatta SWM, et al. energy distribution of positive charges in gate dielectric: Probing technique and impacts of different defects. IEEE Transactions on Electronic Devices. 2013;**60**(5):1745-1753

[38] Huard V, et al. Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs. In: International Reliability Physics Symposium (IRPS); Dallas, Texas, USA, 2003. pp. 178-182

[39] Tan SS, et al. Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations. Applied Physics Letters. 2003;**82**(12):1881-1813

[40] Chakravarthi S, et al. A comprehensive framework for predictive modeling of negative bias temperature instability. In: International Reliability Physics Symposium (IRPS); 2004. pp. 273-282