Maximizing Reliable Performance of Advanced CMOS Circuits—A Case Study

B. Kaczer[1,*], C. Chen[2,†], P. Weckx[1,3], Ph. J. Roussel[1], M. Toledano-Luque[1], J. Franco[1], M. Cho[1], J. Watt[2], K. Chanda[2], G. Groeseneken[1,3], T. Grasser[4]

[1]imec, Belgium; [2]Altera Corp., USA; [3]ESAT KULeuven, Belgium; [4]TU Wien, Austria

[*]e-mail: kaczer@imec.be; [†]e-mail: chrchen@altera.com

***Abstract*—We consider in detail the aspects of maximizing application performance while maintaining its sufficient reliability on the specific case of serially connected nFETs. Serially connected nFETs used in some digital CMOS applications, such as SRAM decoder circuits, and dynamic logic, are vulnerable to Positive Bias Temperature Instability (PBTI). Here we discuss the impact of PBTI frequency and workload on high-k/metal gate nFETs in terms of Capture and Emission Time (CET) maps and quantitatively explain the degradation of our test circuit. This constitutes one of the first validations of the CET map-based methodology on a real silicon circuit. From individual trapping events in deeply scaled nFETs we then project PBTI distributions at 10 years. We further show that at increased supply voltage the serially connected nFET speed degradation is outweighed by signal transfer speedup, resulting in a net performance improvement. Finally, we discuss other degradation mechanisms and conclude the reliability in the studied case will be limited by hard gate oxide breakdown.**

Keywords—reliability; performance; serially connected nFET; circuits; Bias Temperature Instability; Hot Carrier Degradation; Gate Oxide Breakdown

I. INTRODUCTION

Manufacturers frequently aim to maximize the performance of their CMOS applications by various means, notably by increasing supply voltages of critical sub-circuits. This, however, typically occurs at the expense of the application reliability [1]. The question being posed nowadays is "what is the maximum performance we can extract from a given technology while maintaining sufficient reliability?" as opposed to a more conservative "what is the reliability of a given technology at nominal operating conditions?" The task at hand is then finding and guaranteeing the optimum balance between performance and reliability, i.e., *maximizing reliable performance* of CMOS applications.

Specifically, serially connected nFETs (used in SRAM, decoder circuits, dynamic logic [2], etc.) are vulnerable to Positive Bias Temperature Instability (PBTI) [3]. A simple SPICE simulation of serially connected nFETs discussed here shows that the nFET threshold voltage shift ΔV_{th} of ~10 mV causes ~10% stage delay loss at operating voltage (Fig. 1) [4].

We first describe the instability in individual nFET devices in terms of Capture and Emission Time (CET) maps [5] and discuss the frequency and workload dependences within this framework. We then use CET maps to successfully quantitatively explain the degradation of a simple test circuit fabricated in a commercial 28 nm high-k/metal gate technology [4,6]. *To the best of our knowledge this constitutes the first verification of the CET map methodology on a real silicon circuit.*

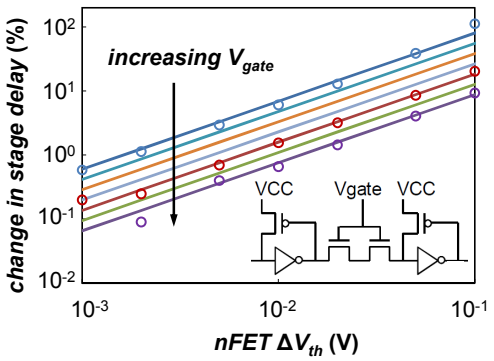

Fig. 1: SPICE-simulated change in average stage delay (symbols; circuit in inset) as a function of the nFET threshold voltage shift ΔV_{th}. The dependence is parameterized (lines). Note that the ΔV_{th} impact on stage delay increases exponentially as V_{gate} is lowered toward V_{CC}, documenting the vulnerability of this topology toward PBTI.

Subsequently, we study individual trapping events in deeply scaled devices of the same technology and we use this information to project time dependent PBTI distributions at 10 years. *Finally, we show that at increased gate voltage V_{gate} the delay degradation of our test circuit is outweighed by signal transfer speedup, resulting in net performance gain at no PBTI penalty.* Moreover, the projected time dependent variability will also decrease at higher V_{gate}. Since SPICE simulations show neither hot carrier degradation nor soft gate oxide breakdown are issues for the serially connected nFETs in our test circuit, the maximum V_{gate} of our test circuit will be limited by hard gate oxide breakdown.

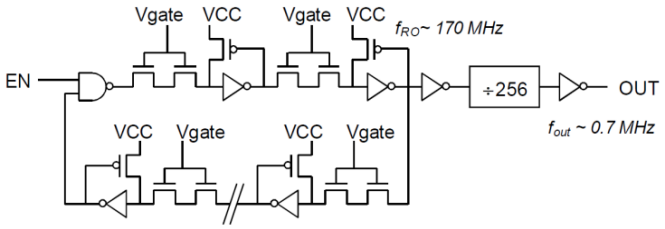

Fig. 2: The ring oscillator (RO) test circuit consists of 127 stages and oscillates at ~170 MHz at operating conditions [6]. Serially connected nFETs are placed between each RO stage and are controlled by a separate terminal V_{gate}. The rest of the circuit is using supply voltage V_{CC}.

TABLE I: SINGLE NFET DEVICES USED IN THIS WORK.

NFET	L (nm)	W (nm)
large	~1000	~1000
short	~28	~1000
small	~28	~100

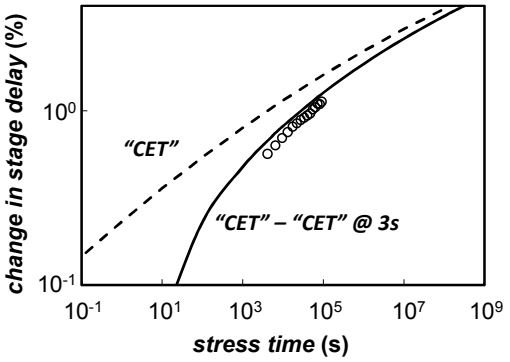

Fig. 3: Change in RO frequency (symbols, normalized to the first reading in 3 s), converted into change in delay per stage, measured at stress V_{gate}, is excellently matched by the projection based on the PBTI CET map (lines).

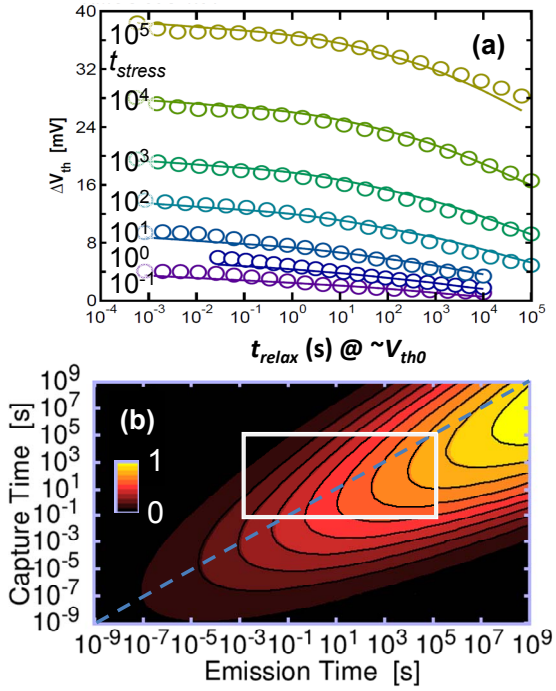

Fig. 4: (a) Threshold voltage shift ΔV_{th} of an individual large nFET (cf. Table I) as a function of stress and relaxation times measured with the eMSM technique [6]. The measurements (symbols) are back-fitted with the CET map in (b) (lines). (b) The data in (a) converted into a CET map (measured data in white rectangle extrapolated using a bivariate Gaussian distribution [5]). The map represents (is proportional to) the probability of finding charged defects with given capture ($\tau_{capture}$) and emission ($\tau_{emission}$) times.

II. EXPERIMENTAL

Individual nFET devices, summarized in Table I, as well as the custom designed test circuit, depicted in Fig. 2 [4,6], were fabricated in the scribe lines of a commercial 28 nm high-k/metal gate technology wafer. The ring oscillator circuit allows simultaneously stressing serially connected nFETs placed between inverter stages by applying stress voltage at V_{gate} and measuring the resulting impact on delay per stage through change in the oscillation frequency.

All measurements were performed at a specified elevated temperature. Because of the limited number of devices per (a large) die, evaluation had to be performed over the entire 300 mm wafer, adding some across-wafer variability to the presented results.

III. RESULTS AND DISCUSSION

The relation between stage delay and threshold voltage shift ΔV_{th} of the serially connected nFET is obtained from SPICE simulations (Fig. 1). Note that the ΔV_{th} impact on stage delay increases exponentially as V_{gate} is lowered toward V_{CC}, documenting the potential vulnerability of the serially connected nFETs toward degradation. At operating conditions the degradation of a mere 10 mV in V_{th} can result in a significant delay degradation of the order of 10%.

The circuit in Fig. 2 is used to measure the change in the stage delay at stress V_{gate} and nominal V_{CC} in Fig. 3. We now show that this measurement can be excellently *quantitatively* explained by PBTI in the serially connected nFET. PBTI is discussed entirely in terms of the corresponding CET maps.

A. PBTI in CET map

The CET map allows describing the entire PBTI degradation process in terms of the probability density of capture and emission times of the responsible gate oxide defects [5]. The map, as extracted from an extended Measure-Stress-Measure (eMSM) sequence [7] on a single nFET device (Fig. 4a), is agnostic of the underlying physical model [8]. A large device has been used to avoid variability due to single-carrier events [9-12]. The "raw" map data in the measurement window were fitted with a single bivariate Gaussian distribution (Fig. 4b), allowing both smoothing and extrapolating from the measurement window towards both very short (clock cycle) and long (application lifetime) times [5].

Fig. 5a illustrates how the CET map is used to construct the time dependence of DC PBTI degradation of a single nFET. Only traps in the rectangle delimited by $\tau_{capture} \in [0, t_{stress}]$ and $\tau_{emission} \in [t_{relax}, \infty]$ contribute to the device's ΔV_{th}. A simple integration thus reproduces the original data in Fig. 4a as well as shorter eMSM measurements performed at different overdrive voltages $V_{ov} = V_{gate} - V_{th}$ (Fig. 5b). Note in Fig. 5b that the projection based on the CET map is not a perfect power law [5], which in fact reproduces the same trend in the measured ΔV_{th} data. It also results in a more optimistic projection of degradation at 10 years.

Fig. 6 illustrates how the same CET map describes the *AC* PBTI degradation; specifically it shows which traps are charged (occupied) traps and contributing to ΔV_{th} [13]. This is

possible provided PBTI process is first order (i.e., traps exist in a single charged state and a single discharged state) [5]. Integration over the occupied traps gives ΔV_{th} independent of high frequencies, which is also reproduced experimentally on large devices (Fig. 7a) [14]. The most revealing confirmation of the CET map-based approach is however in Fig. 7b—the experimentally observed workload dependence is excellently reproduced.

Note that, in contrast to NBTI [15], the degradation at DF = 50% is ~90% of the DC (i.e., DF = 100%) value. This is because the PBTI traps are slow to discharge (emit), as documented in Fig. 4b—the peak maximum is below the $\tau_{capture} = \tau_{emission}$ line.

The serially connected nFETs in the test circuit in Fig. 2 are, however, based on *short* NFET devices. Fig. 8 documents that short devices degrade less in this technology. (Note that the short devices also have a slightly lower V_{th0}, not shown) This is likely due to a thicker gate oxide (and the resulting lower electric field), as discussed e.g. in [16].

PBTI frequency independence is confirmed in short devices in Fig. 7a, allowing us to apply the same CET approach to scale short device PBTI degradation to ~100 MHz and excellently quantitatively reproduce the measured change in stage delay, see Fig. 2. *The CET map-based methodology is thus validated in real silicon circuits.*

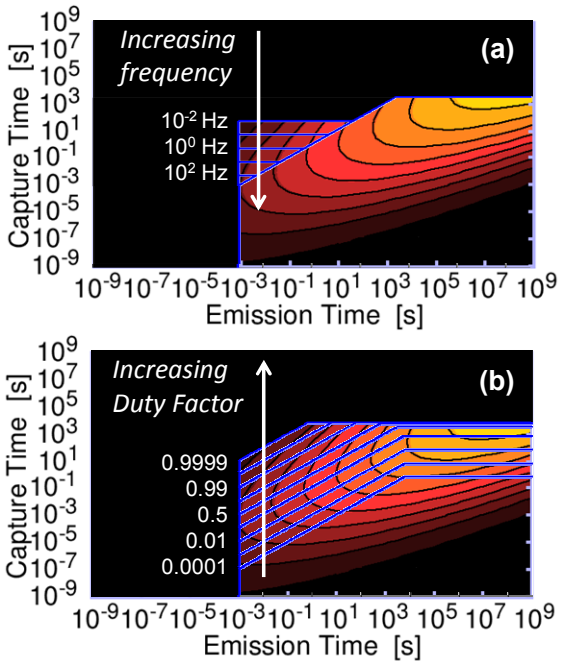

Fig. 6: Analogously to Fig. 5a, the CET map of Fig. 4b overlaid with an occupancy map allows representing stress at (a) different frequencies and (b) duty factors. (a) A weak frequency f dependence is expected at low frequencies [9]. At higher frequencies the occupancy tends toward beveled rectangle delimited by t_{stress} and t_{relax}. (b) The beveled-rectangle region increases with increasing Duty Factor (f = 1 kHz).

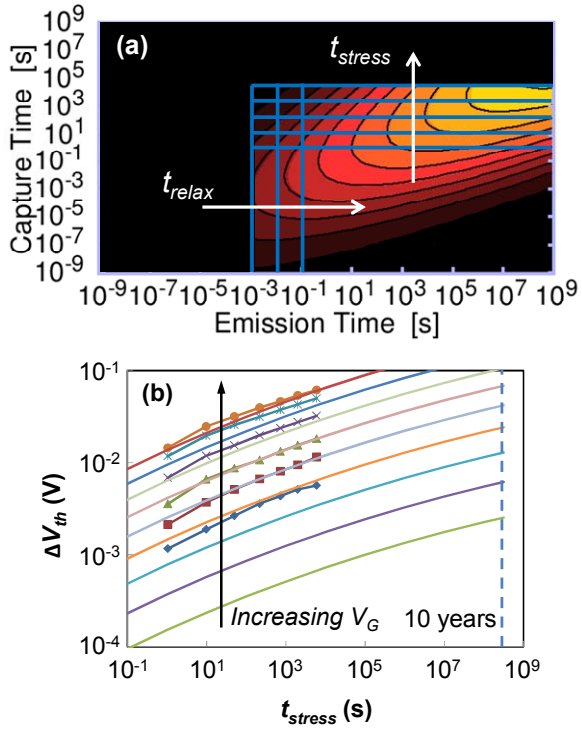

Fig. 5: (a) The CET map of Fig. 4b overlaid with an occupancy map representing a shorter, DC stress eMSM measurement. Only traps in the region delineated by blue vertical and horizontal lines are occupied and thus are contributing to ΔV_{th}. The largest rectangle corresponds to the maximum stress time (6,000 s, cf. (b)) and the shortest relaxation time (~1 ms) used in the measurement. (b) Measured ΔV_{th} at t_{relax} = 1 ms as a function of stress times at different stress overdrives (symbols) excellently described by integrating the CET map in (a) (lines). Voltage scaling is done as per Fig. 8.

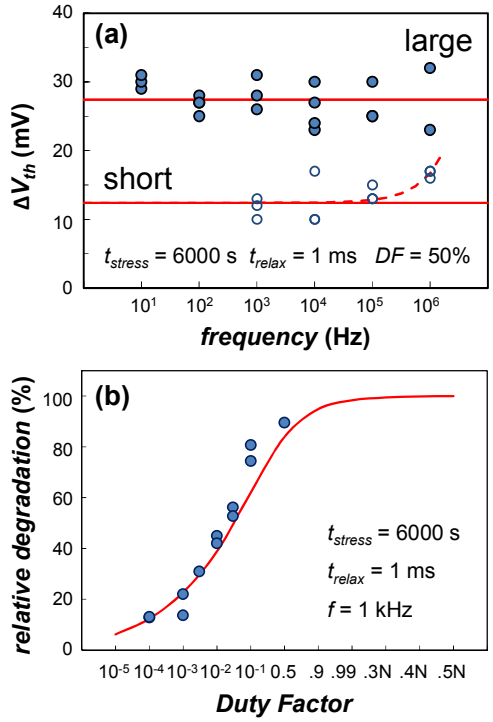

Fig. 7: (a) Frequency-independent threshold voltage shifts and (b) the relative decrease of ΔV_{th} at low duty factors are well predicted by integrating the corresponding areas shown in Figs. 6a and 6b, respectively (note: the stress time is the duration of application of AC signal). Solid circles: device data, solid lines: CET map integration. Dashed line: A possible increase at higher frequencies f can be traced to an overshoot observed with a digital storage oscilloscope in the leading edge of V_{gate} AC waveform (not shown), thus $\Delta V_{th} = \Delta V_{th,low\text{-}f} + a f$.

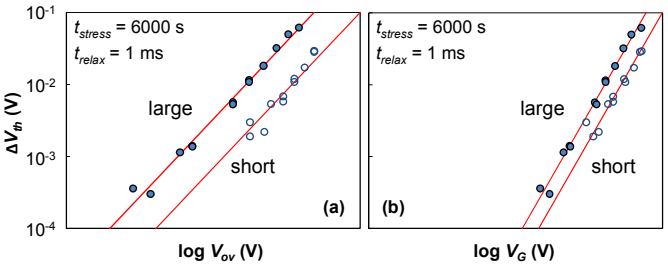

Fig. 8: (a) Voltage overdrive V_{ov} and (b) gate voltage V_G acceleration of PBTI degradation of large and short devices. Short devices degrade less, likely due to thicker gate oxide.

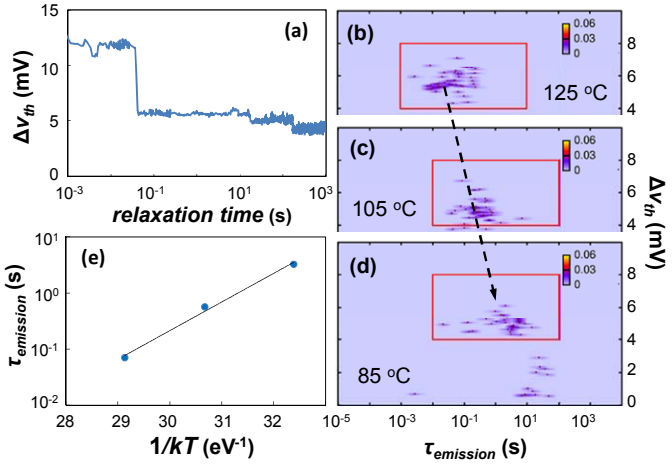

Fig. 9: (a) Illustration of a single relaxation transient of a single small nFET. (b-d) TDDS: A selected single electron trap emission (note the "shifting" with T of defect cluster due to multiple perturbations/ transients) is strongly thermally activated, E_a = 1.18 eV in this particular case (e) [12].

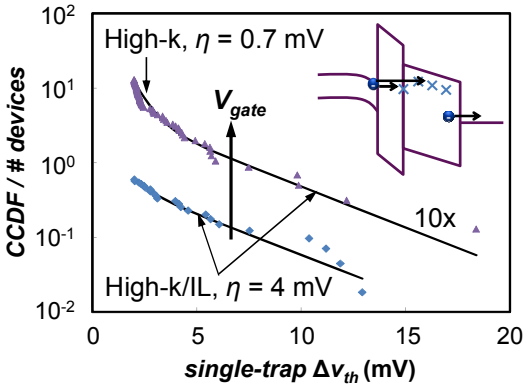

Fig. 10: Distributions of V_{th} shifts due to single trapping events (collected from transients as in Fig. 9a). The bimodal distribution corresponds to trapping in high-k and at IL/high-k interface [12] (inset; the infrequently-observed electron de-trapping also illustrated). At lower V_{gate}, trapping at the IL/high-k interface is dominant. Both small nFET η values indicated in the figure are also consistent with η ~ 6mV extracted on small *pFETs* (not shown), corresponding to hole trapping at Si/IL interface.

B. *Reliability and performance projections*

In order to find the optimum balance between reliability and performance of the serially connected nFETs, we now consider both these aspects as a function the gate voltage V_{gate}. We first show that the PBTI reliability (both the mean and the time-dependent variability) of serially connected nFETs will actually improve with increasing V_{gate}. We then discuss other degradation mechanisms, namely Channel Hot Carrier Degradation (HCD) and Soft and Hard gate oxide breakdowns (SBD and HBD), and conclude that the reliability will be limited by HBD.

B1. BTI reliability and performance

To assess the time dependent variability of serially connected nFETs, we study the properties and impact of individual (charged) gate oxide traps. We then combine this information with mean PBTI degradation projections to generate the expected ΔV_{th} distributions [9,11].

Single electron trap discharge events were studied in the *small* devices in Time-Dependent Defect Spectroscopy (TDDS) experiments (Fig. 9) [10]. Consistent with the distributed activation energies underlying the CET maps, strong temperature activation of individual defects has been observed (Figs. 9b-e) [11,12]. As in [11], a bimodal exponential distribution of single electron ΔV_{th}'s was observed with expectation values η of 0.7 and 4 mV, respectively, as is evidenced in Fig. 10. The following time dependent variability analysis was based on the latter, i.e., the wider distribution, since these steps were persistent at lower stress voltages.

The knowledge of η and the mean $\langle \Delta V_{th} \rangle$ from the CET map allows us to project distributions of ΔV_{th} (hence the delays, cf. Fig. 1) of the serially connected nFETs in a large area circuit at 10 years at the specified T and varying operating V_{gate} using

$$H_{\eta,N_T}(\Delta V_{th}) = \sum_{n=0}^{\infty} \frac{e^{-N_T} N_T^n}{n!} \left[1 - \frac{n}{n!} \Gamma(n, \Delta V_{th} / \eta) \right], \quad (1)$$

where $N_T = \langle \Delta V_{th} \rangle / \eta$ [9]. The mean $\langle \Delta V_{th} \rangle$ values, projected as in Fig. 5b (for the short devices; not shown), and the $\pm 3\sigma$ values are plotted vs. varying V_{gate} in Fig. 11a. The figure illustrates that although the nFETs are expected to shift little *on average*, a non-negligible fraction of devices is expected to shift multiples of the mean value. This time dependent variability thus needs to be taken into consideration already during the application design phase.

The mean and the $\pm 3\sigma$ ΔV_{th} values from Fig. 11a were converted into the change in stage delay (Fig. 1) and replotted vs. varying V_{gate} in Fig. 11b. As expected (cf. Fig. 8), higher stress of the serially connected nFET results in more severe degradation of the stage delay. However, higher V_{gate} also results in a faster operation of each stage, as calculated with SPICE and also shown in Fig. 11b. The overall change in serially connected nFET delay will be the product of these two effects, shown in Fig. 11c. *The figure shows that the speedup due to increased V_{gate} outweighs the PBTI degradation* (cf. Table II of [6]). Moreover, the projected time dependent variability will also decrease at higher V_{gate}. We therefore conclude the performance of the serially connected nFET can be increased without any PBTI penalty.

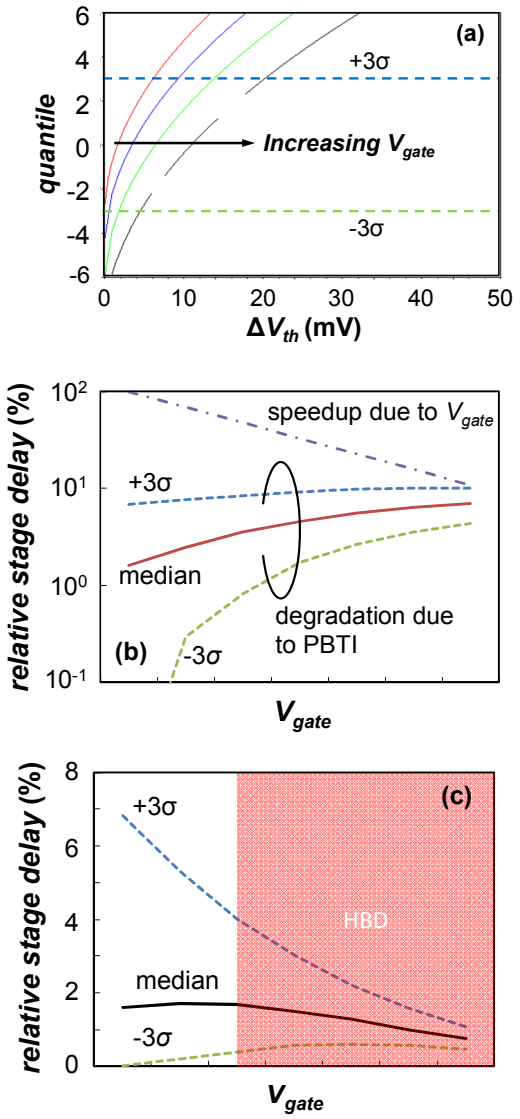

Fig. 11: (a) The threshold voltage shifts V_{th} in individual serially connected nFETs in an actual application will be distributed. The distributions at varying V_{gate}'s are based on 10 year projected mean value $\langle \Delta V_{th} \rangle$, η = 4 mV, and Eq. 1. (b) The evolution of distributions in (a) described through the ±3 σ and mean depenences on V_{gate}. The threshold voltage shifts are converted into relative stage delays through Fig. 1b. Also shown is the relative stage speedup due to increased V_{gate}. (c) With increasing V_{gate}, the relative stage delay will degrade due to increased BTI but will simultaneously improve due to speedup (cf. (b)). The overall effect of increasing V_{gate} will thus be positive, while, at the same time, the variability will decrease. Maximum V_{gate} will be limited by HBD (cf. Fig. 15).

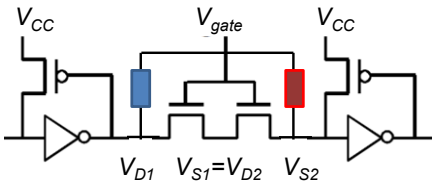

Fig. 12: Node and additional breakdown leakage path definitions used in our SPICE simulations.

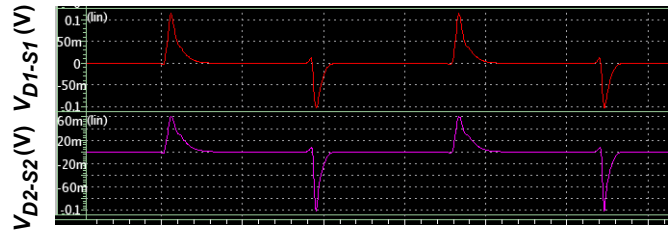

Fig. 13: Circuit simulations show the serially connected nFETs will not be subjected to hot carrier degradation conditions (max $V_{d\text{-}s}$ bias ~ 0.1 V, node definitions cf. Fig. 12).

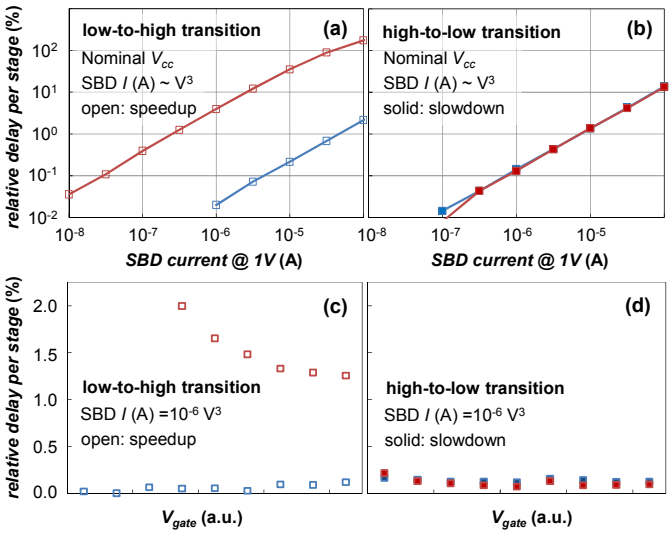

Fig. 14: (a-b) Circuit simulations show that a non-ohmic SBD (red/blue BD location definition cf. Fig. 12) of the expected magnitude < 1 µA @ 1V will have a limited impact on circuit delay (in fact, speedup for low-to-high delay is expected (a)). (c-d) Although increased V_{gate} will accelerate BD, the impact of SBD on the delay will *not* be higher at higher V_{gate}.

B2. Other degradation mechanisms

We now discuss other degradation mechanisms potentially affecting the serially connected nFETs, namely HCD, SBD and HBD. Circuit definitions aiding this discussion are given in Fig. 12. SPICE simulations in Fig. 13 show that the serially connected nFET will never be subjected to strong source-drain biases, thus excluding any significant HCD degradation.

SBD is the typical mode of first breakdown in the high-k dielectric such as that used in our devices, with typical magnitudes below 1 µA at stress conditions. Since the severity (magnitude) of the BD decreases strongly with voltage [17], we expect SBD occurring at operating voltage to have equivalent leakage at 1V far below 1 µA (note that due to the SBD non-linearity the leakage has to be specified at a given voltage). SPICE simulations in Fig. 15 show that SBD affects the speed of the serially connected nFET. Due to the "pre-charging" effect of the SBD leakage path [17], the low-to-high transition delay of the serially connected nFET can be in fact improved (Fig. 14a), albeit always at the cost of

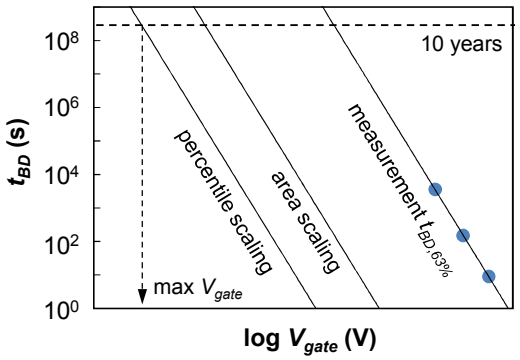

Fig. 15: HBD projection determines the maximum V_{gate} (see also Fig. 11c). Analytical technique to reduce across-wafer variability [19] and the total gate area of all sub-circuitry exposed to higher V_{gate} are used.

increased power consumption. The case in which the performance (speed) is actually degraded is shown in Fig. 14b. This figure, however, documents that degradation below 0.1% is expected for SBDs with equivalent 1 µA at 1V. Figs. 14c and 14d then illustrate a somewhat counterintuitive observation that the SBD, although more conductive at higher V_{gate}, will not impact the serially connected nFET speed more severely.

We therefore conclude that the maximum V_{gate} will be limited by HBD. HBDs, with typical equivalent resistances of 1-10 kΩ, are expected to completely disable all but the most conservatively scaled serially connected nFETs [18], as can be also inferred Fig. 14 (assuming currents 0.1 – 1 mA expected for HBD). The projection in Fig. 15 enables us to set the maximum limit on V_{gate} (cf. Fig. 11c), *allowing us to extract maximum performance from the given technology while maintaining sufficient reliability of the application.*

IV. Conclusions

We have considered in detail the aspects of maximizing performance while maintaining sufficient reliability on the specific case of serially connected nFET. PBTI frequency and workload dependences in high-k/metal gate nFETs have been discussed within the CET map framework, allowing us to quantitatively explain the degradation of our test circuit. To the best of our knowledge this constitutes the first validation of the CET map-based methodology up to the real silicon circuit level. Furthermore, we have shown the performance of this particular circuit topology can be increased at no PBTI reliability penalty. We suggest that its reliability will be limited by HBD.

Acknowledgments

The presented work has been performed within the imec INSITE Partner Program. Detailed results were accessible to all imec Core and imec INSITE members 6 months prior to publication.

References

[1] T. Nigam, "Scaling to the Final Frontier: Reliability challenges in sub 20 nm technologies", Proc. IEEE Int. Integrated Reliab. Workshop (IIRW), p. xi, 2011.

[2] B. Kaczer and G. Groeseneken, "Potential vulnerability of dynamic CMOS logic to soft gate oxide breakdown", IEEE Electron Dev. Lett. 24(12), pp. 742–744, 2003.

[3] R. Degraeve *et al.*, "Review of Reliability Issues in High-k/Metal Gate Stacks", 15th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), pp. 1–6, 2008.

[4] B. Kaczer *et al.*, "Reliability and Performance Considerations for NFET Pass Gates in FPGA Applications", Proc. Int. Integrated Reliab. Workshop (IIRW), 2013.

[5] T. Grasser *et al.*, "Analytic modeling of the bias temperature instability using capture/emission time maps", Int. Electron. Dev. Meeting (IEDM) Tech. Dig., pp. 27.4.1–27.4.4, 2011.

[6] C. S. Chen and J. T. Watt, "Characterization and simulation of NMOS pass transistor reliability for FPGA routing circuits", IEEE Int. Conf. Microel. Test Structures (ICMTS), pp. 216–220, 2013.

[7] B. Kaczer *et al.*, "Ubiquitous Relaxation in BTI Stressing—New Evaluation and Insights", Proc. Int. Reliab. Phys. Symp. (IRPS), pp. 20–27, 2008.

[8] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The Statistical Analysis of Individual Defects Constituting NBTI and its Implications for Modeling DC- and AC-Stress," Proc. Int. Reliab. Phys. Symp. (IRPS), pp. 7–15, 2010.

[9] B. Kaczer, Ph. J. Roussel, T. Grasser, and G. Groeseneken, "Statistics of Multiple Trapped Charges in the Gate Oxide of Deeply Scaled MOSFET Devices—Application to NBTI", IEEE Electron Dev. Lett. 31(5), pp. 411–413, 2010.

[10] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability", Proc. Int. Reliab. Phys. Symp. (IRPS), pp. 16–25, 2010.

[11] M. Toledano-Luque *et al.*, "From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation", Proc. VLSI Symp., p. 152–153, 2011.

[12] M. Toledano-Luque *et al.*, "Temperature and voltage dependences of the capture and emission times of individual traps in high-k dielectrics", Microelectronic Eng. 88(7), pp. 1243–1246, 2011.

[13] J. Martin-Martinez *et al.*, "Probabilistic defect occupancy model for NBTI", Proc. Int. Reliab. Phys. Symp (IRPS)., pp. XT.4.1–XT.4.6, 2011.

[14] K. Zhao, J. H. Stathis, B. P. Linder, and E. Cartier, "PBTI under dynamic stress: From a single defect point of view", Proc. Int. Reliab. Phys. Symp. (IRPS), pp. 4A.3.1–4A.3.9, 2011.

[15] T. Grasser, B. Kaczer, H. Reisinger, P.-J. Wagner, M. Toledano-Luque, "On the frequency dependence of the bias temperature instability", Proc. Int. Reliab. Phys. Symp. (IRPS), pp. XT. 8.1–XT. 8.7, 2012.

[16] J. C. Liao *et al.*, "Strain effect and channel length dependence of bias temperature instability on complementary metal-oxide-semiconductor field effect transistors with high-k/SiO_2 gate stacks", Appl. Phys. Lett. **93**, p. 092101, 2008.

[17] B. Kaczer, R. Degraeve, R. O'Connor, Ph. Roussel, and G. Groeseneken, "Implications of progressive wear-out for lifetime extrapolation of ultra-thin (EOT ~ 1 nm) SiON films", Int. Electron. Dev. Meeting (IEDM) Tech. Dig., pp. 713–716, 2004.

[18] B. Kaczer *et al.*, "Analysis and modeling of a digital CMOS circuit operation and reliability after gate oxide breakdown: a case study", Microelectronic Reliab. 42 (4-5), pp. 555–564, 2003.

[19] Ph. J. Roussel., R. Degraeve, A. Kerber, L. Pantisano, and G. Groeseneken, "Accurate reliability evaluation of non-uniform ultrathin oxynitride and high-k layers", Proc. Int. Reliab. Phys. Symp. (IRPS), pp. 29–33, 2003.