

NBTI DEGRADATION: FROM TRANSISTOR TO SRAM ARRAYS

V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes and L. Camus
STMicroelectronics, Crolles, France
phone: +33438922907; fax: +33438923227; e-mail: Vincent.huard@st.com

PURPOSE

A novel composite model had been recently introduced to physically explain the mean pMOS threshold voltage shift (V_{TP}) induced by NBTI degradation at transistor level in a quantitative way. This model is here extended to include the statistical variations introduced by intrinsic fluctuations. In a second time, the model is extrapolated up to SRAM arrays by analyzing the SRAM bitcell sensitivity to transistor degradation. This approach allows quantitative prediction of NBTI-induced V_{MIN} variations and access time T_{aa} degradation during burn-in operations. The key findings include (a) demonstration of non-normality of V_{TP} shift distribution (b) NBTI contribution to product V_{MIN} drift arises from both mean V_{TP} drift but also from increased V_{TP} dispersion, and (c) V_{TP} shift non-normality is smoothed out at product level by time-zero variation of the six transistors of the SRAM bitcell. [Keywords: NBTI, composite model, SRAM, Static Noise Margin, Write Margin, Access time, Skellam.]

INTRODUCTION

In advanced technology nodes, both variability and reliability are important barriers to Static Random Access Memory (SRAM) bitcell scaling [1-2]. The resulting reduction of transistor geometries has caused stronger inter-die variations in process parameters (such as V_T , L , W ...) and intra-die variations in the number and spatial distribution of dopants in the channel region. Among them, the random placement of dopants is more of a concern since it cannot be eliminated by a tight control of the manufacturing process. In addition, NBTI degradation introduces systematic V_{TP} drift in correlation with increased intra-die variability. As a consequence, it is necessary to optimize not only a SRAM cell for time zero yield but also for reliability degradation.

In this paper, the mean NBTI degradation is modeled using the newly developed composite model. This model is here extended to deal with the additional statistical variations introduced by NBTI degradation, which is of prime importance when studying SRAM arrays. In a second time, the SRAM bitcell sensitivity to NBTI degradation is both simulated using in-house latest Design-in Reliability tools and experimentally checked. Finally, device and bitcell parameters are quantitatively linked to SRAM new metrics, called F_{bit} , i.e. the number of faulty cells for a given number of bitcells. This new metric will be shown to allow quantitative optimization to achieve both product yield and robust cells against NBTI drift. In addition to V_{MIN} drift, NBTI degradation will also slow down the time delay from the clock edge to valid data presented on the bus (T_{aa}). Product level testing has been done on both ST 65nm and 45nm Low Power (LP) platforms to optimize SRAM bitcells through quantitative correlation with simulations.

MEAN NBTI DEGRADATION: TRANSISTOR LEVEL

The way NBTI degradation is modeled at transistor level has a great impact on circuit predictions. For many years, Reaction-Diffusion (RD) model had been the dominant modeling approach for the median V_{TP} shift [3]. In this paper, we use the novel composite approach combining both a permanent part D_p , which consists of interface traps created during stress, and a recoverable part D_R , which

consists of valence hole trapping/detrapping in preexistent traps [4]. Let's here summarize the main conclusions of the composite model.

Permanent part: Interface traps

The interface traps are permanently created (i.e. recovery is decades longer than the transient effect commonly observed). A reaction-limited (RL) model for interface traps creation has been proposed which predicts both the linear temperature dependence of the power law time exponent (Fig.1), already reported in [5-6], and saturation effects when all available bonds are broken.

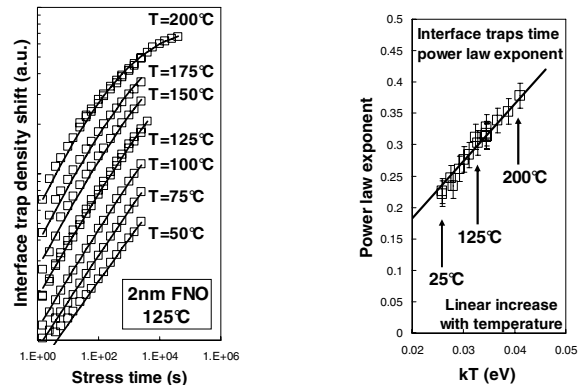


Figure 1: Reaction-Limited (RL) Model (lines) vs experiments (squares) for various temperatures (Left part). Linear dependence of power law time exponent is described by RL model (right part).

Recoverable part: Hole trapping/detrapping

Though the permanent part is dominating digital products degradation at long-term, the recoverable part will dominate the frequency and transistor activity dependences. Hole trapping/detrapping explanation for the recoverable part is favored, as pointed out by the low experimental temperature activation. In this framework, both valence band and interface states should contribute as pathways to fill up the oxide traps during stress phase (i.e. Fermi level pinned in valence band) [10]. The filling dynamics is characterized by the time constant τ_{otc} . The existence of these two pathways can be experimentally evidenced during recovery phase, characterized by the time constant τ_{ote} . If the Fermi level remains pinned in valence band during the recovery (i.e. recovery voltage remains above V_{TP}) only valence band states should contribute to recovery (Fig. 2). Experimental recovery dynamics are identical for two different recovery voltages above V_{TP} (Fig. 3). If the Fermi level lies within the silicon gap, an additional pathway of detrapping is triggered through the interface states. This additional pathway is evidenced through an accelerated recovery when the Fermi level is moved from valence band (-0.35V) to conduction band (+1V) (Fig. 3). Another typical characteristic of detrapping phenomenon is the universality of recovery dynamics for given stress conditions versus the ratio stress time over recovery time (t_s/t_r) (Fig. 4).

Filling $|V_{g1}| > |V_t|$ vb & it states $\tau_{otc} \gg \tau_{it}$

Emptying $|V_{g2}| < |V_{g1}|$ $|V_{g2}| > |V_t|$: vb states

$|V_{g2}| < |V_t|$: vb & it states

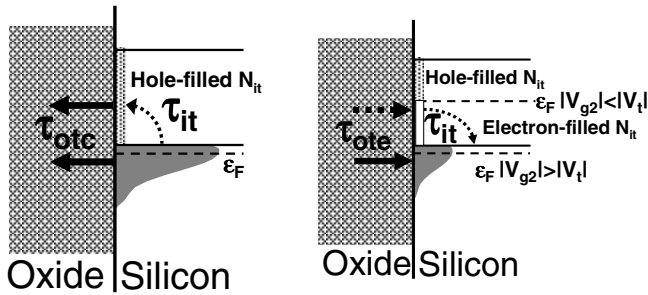


Figure 2: Hole trapping/detrapping scheme through both valence and/or interface states.

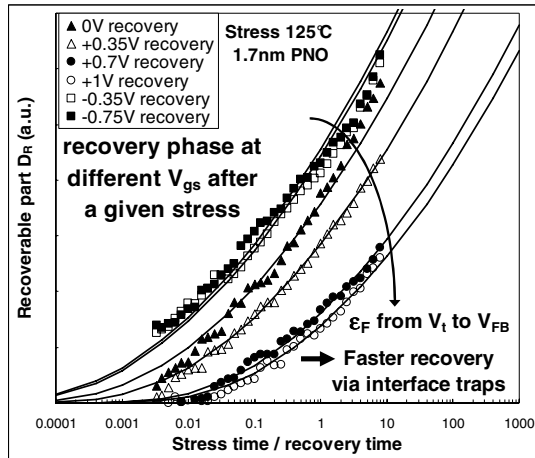


Figure 3: Experimental recovery dynamics (symbols) for recovery voltages ranging from V_{TP} to V_{FB} , for plasma-nitrided oxides (PNO), compared to composite model predictions (lines).

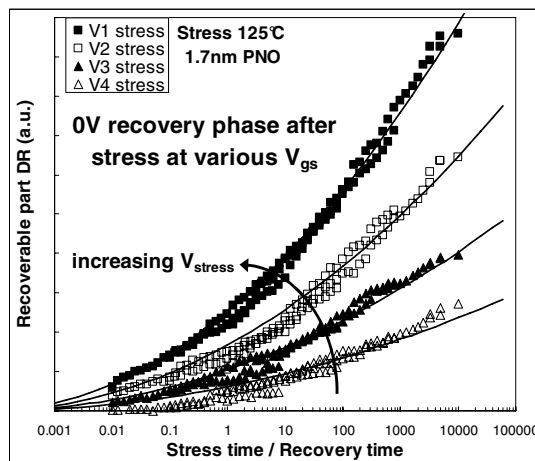


Figure 4: Experimental recovery dynamics (symbols) following stress phases with various stress voltages, compared to composite model predictions (lines).

Long term predictions: comparison with RD model

It is important here to focus on the impact of model choice in terms of long-term product predictions. As mentioned previously [7-8], RD model successfully predicts stress phase but fails to cover accurately the recovery phase inherent to any products. Figure 5 shows that our composite approach allows an accurate modeling of recovery phase compared to RD model, with similar time dynamics for stress phase.

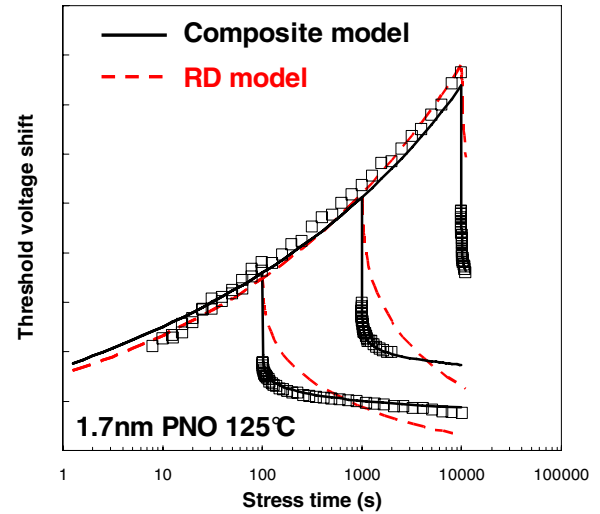


Figure 5: Composite NBTI (full line) and RD (dashed line) models benchmarked vs experiments for both stress and recovery phases.

Accurate modeling of transient effects is of prime importance for product lifetime predictions. First, since most of the products are working under AC conditions, the average degradation strongly depends on the average transistor activity including stress frequency and/or duty cycle. Figure 6 shows that both models predict frequency independence of NBTI degradation, but the different recovery dynamics yield to various degradation levels for a given duty cycle. In a second step, accurate recovery time dynamics modeling is also needed for relevant duty cycle description for a given frequency, as shown in bottom part of figure 6, where discrepancies between both models are revealed. Our novel composite model clearly allows a more accurate description of the mean NBTI degradation of both quasi-static and AC results.

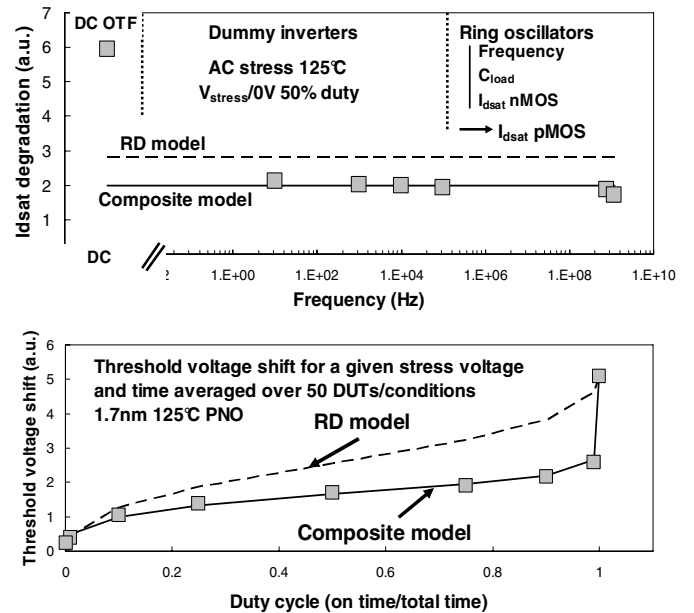


Figure 6: Composite NBTI (full line) and RD (dashed line) models benchmarked vs experiments (squares) for various frequencies (top part) (for 50% duty stress) and for various duty cycles (bottom part) (for 100kHz). In top part, non recovered DC degradation (DC OTF) is mentioned for reference purpose.

STATISTICAL NBTI DEGRADATION

The composite model proposes an accurate description of the mean V_{TP} degradation of pMOSFETs in SRAM bitcells as a function of cell activity. Nevertheless, modeling the impact of NBTI degradation for an assembly of SRAM bitcells, it is important to deal with the statistical dispersion of V_{TP} shift. These intrinsic fluctuations are coming from the fact that since the NBTI degradation is caused by discrete charges, either interface traps generation or hole trapping, the underlying stochastic process is discrete. The discrete nature of NBTI-generated charges can be revealed by tracking V_{TP} shift recovery in SRAM-sized transistor (Figure 7) [9].

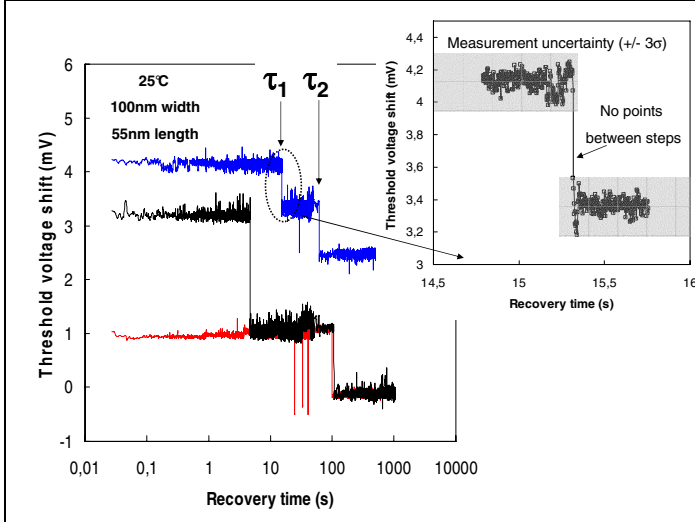


Figure 7: Recovery of V_{TP} shifts for several SRAM-sized devices showing the discrete nature of detrapping events. The inset shows that there are no intermediary values between steps in spite of the small measurement step time (shaded area represents the measurement uncertainty $(\pm 3\sigma)$).

V_{TP} shift variance: Normal approximation

We first focus on the generation of interface traps, i.e. the permanent part. The total number of charges ΔN induced by NBTI stress is directly related to V_{TP} shift by:

$$\Delta V_{TP} \propto \frac{q}{C_{ox_INV} \cdot A_G} \cdot \Delta N = K_N \cdot \Delta N \quad \text{eqn. 1}$$

where A_G is the effective gate area.

It is commonly assumed that the random fluctuations in the number of induced charges ΔN follow a Poisson distribution, so we can write:

$$\begin{aligned} \text{Var}(\Delta N) &= \text{Mean}(\Delta N) \\ &\propto \frac{1}{K_N} \cdot \text{Mean}(\Delta V_{TP}) \quad \text{eqn. 2} \end{aligned}$$

which finally relates to the variance of the induced V_{TP} shift as:

$$\text{Var}(\Delta V_{TP}) = K \cdot K_N \cdot \text{Mean}(\Delta V_{TP}) \quad \text{eqn. 3}$$

with $K=1$, since only the random fluctuations of the number of induced charges are taken into account. Nevertheless, it was shown that the random fluctuations in the spatial distribution of induced charges contribute to V_{TP} shift variance and is of roughly the same magnitude than number fluctuations [9]. To accommodate this additional variation, the empirical constant K is introduced with an expected value of 2.

Figure 8 illustrates that V_{TP} shift distributions of the permanent part can roughly be estimated by normal distributions with an increasing variance when mean V_{TP} shift increases, as expected from eqn. 3.

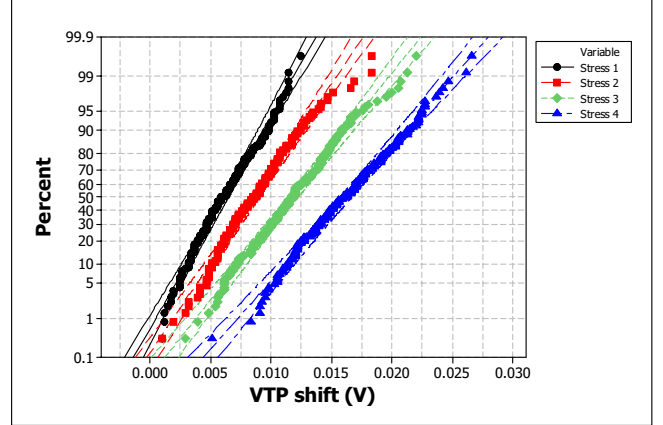


Figure 8: V_{TP} shift distributions for interface traps creation for various stress times showing the increased dispersion when V_{TP} shift increases.

The correlation plot of the Variance (V_{TP} shift) to the Mean (V_{TP} shift) allows determining accurately the experimental constant K (Figure 9), which is equal to 2, in agreement with theoretical value [10]. At this point, it is also important to notice that the variance of V_{TP} shift does not depend on stress conditions, i.e. eqn. 3 is valid whatever the stress conditions (Figure 9 where different symbols representing different stress voltages).

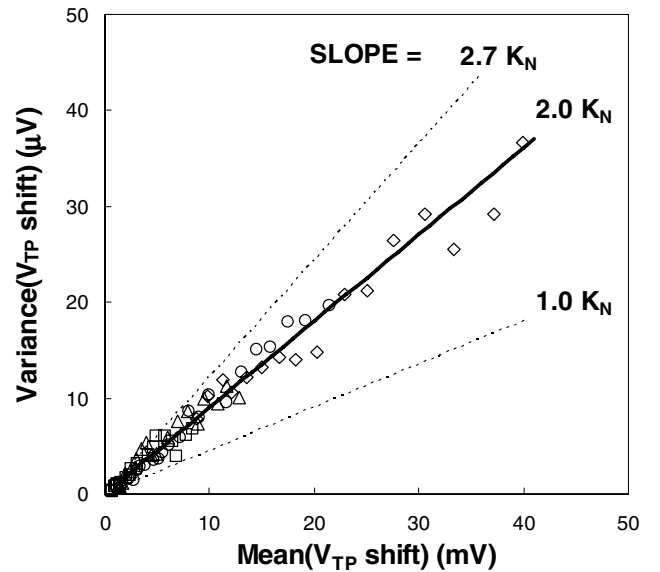


Figure 9: V_{TP} shift variance is fully described by mean V_{TP} shift, independently to the stress conditions (five different voltages in this case represented by different symbols), and in agreement with eqn. 3 (Solid line). The K factor value is found equal to 2 in agreement with theoretical prediction.

When dealing with SRAM arrays with millions of bitcells, the Normal approximation of the V_{TP} shift distribution needs to get examined more rigorously from a statistical point of view. V_{TP} shift distributions have been monitored on SRAM-sized transistor from our 45nm process flow with a large sampling of more than a thousand transistors. Slow V_{GS} sweeps with long waiting time at each readout time have been used in this study to focus on the permanent part (i.e. interface traps creation). First, we have checked that the variance of V_{TP} mismatch shift is two times larger than the variance of V_{TP} shift. This result confirms that the variation in V_{TP} shift on small dimensions transistors is dominated by random fluctuations over process variations. Figure 10 shows the measured V_{TP} shift distributions for three different readout times. Normal distribution approximation with the same mean and variance underestimates the high V_{TP} shifts tails of the actual experimental distributions.

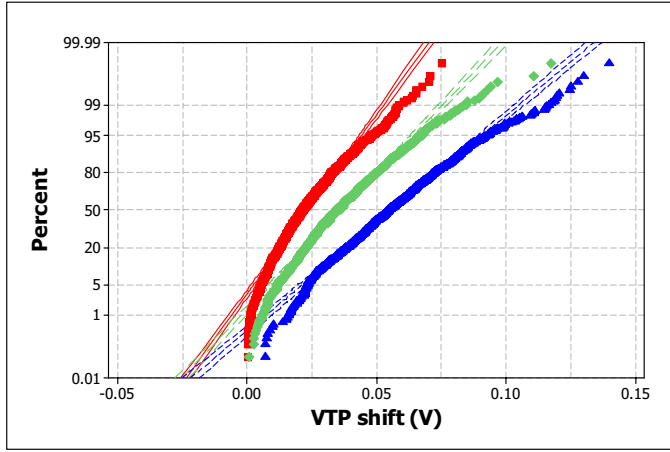


Figure 10: Experimental V_{TP} shift distributions (symbols) for SRAM-sized transistors measured at three different readout times. Normal distribution approximation (dotted lines with 95% confidence bounds) with the same mean and variance underestimates the high V_{TP} shifts tails of the actual experimental distributions.

More interestingly, the corresponding “fresh” and “degraded” V_{TP} distributions follow Poisson distributions (Figure 11).

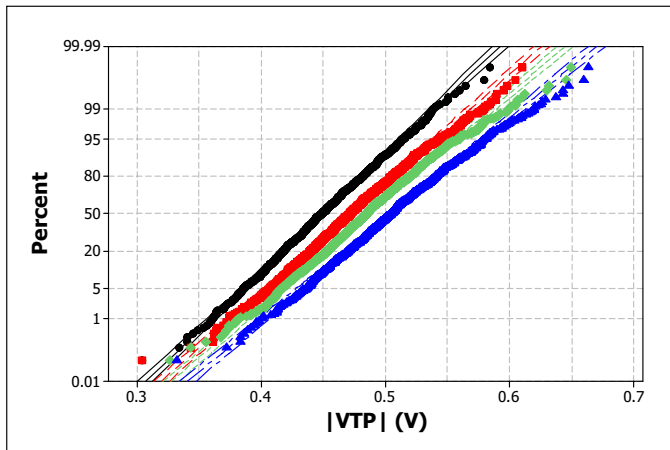


Figure 11: Experimental V_{TP} distributions (symbols) for SRAM-sized transistors measured at time zero (black circles) and three different readout times. Experimental V_{TP} distributions obey Normal distribution approximation (dotted lines with 95% confidence bounds) with the same mean and variance.

This apparent discrepancy was solved out decades ago by Skellam [11] who first demonstrated that the difference between two Poisson distributions is not Poisson. This approach was recently reexamined by Rauch [12] who considers that V_{TP} shift is resulting from the difference C-D of two uncorrelated random variables C and D (for “creation” and “destruction”), both Poisson distributed with means μ_C and μ_D . The Creation process is in this case related to interface traps creation and the Destruction process to the repassivation process (a prominent feature of the RD model). It was demonstrated that, for U a normalized random variable with a mean of 1, it is possible to write:

$$\begin{aligned}\mu(\Delta V_{TP}) &= K_N \cdot (\mu_C - \mu_D) \\ \text{Var}(\Delta V_{TP}) &= K_N^2 \cdot (\mu_C + \mu_D) \cdot (1 + \text{Var}(U))\end{aligned}\quad \text{eqn. 4}$$

which allows defining the dispersion factor ϕ as

$$\begin{aligned}\phi &= \frac{\text{Var}(\Delta V_{TP})}{K_N \cdot \mu(\Delta V_{TP})} = \frac{(\mu_C + \mu_D)}{(\mu_C - \mu_D)} \cdot (1 + \text{Var}(U)) \\ \text{Var}(U) &\approx 0.8 - 1.1\end{aligned}\quad \text{eqn. 5}$$

ϕ represents the “overdispersion” compared to what would result from a single Poisson process. Assuming an asymptotic approximation to the compound distribution for large N, the corresponding cumulative distribution function (CDF) is [12]:

$$F(\Delta V_{TP}) \approx \sum_{m=0}^{\infty} \Phi\left(\frac{\Delta - m}{\sqrt{m \cdot (\phi - 1)}}\right) \cdot \frac{e^{-\mu(\Delta)} \mu(\Delta)^m}{m!} \quad \text{eqn. 6}$$

where $\Delta = \Delta V_{TP}/K_N$ and Φ is the Standard Normal CDF. For $m=0$, Φ is taken to be the unit step function $H(x)$. The comparison of experimental CDFs with results from eqn. 6 (Figure 12) shows excellent agreement using $\phi=2$, as extracted from Figure 9.

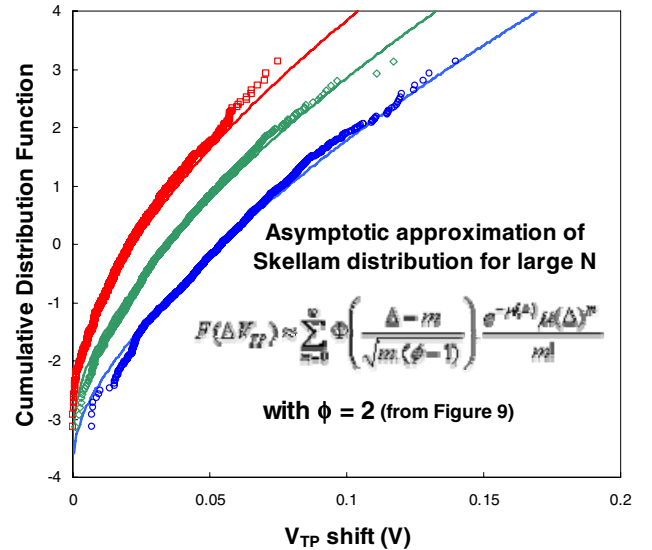


Figure 12: Comparison of experimental V_{TP} shift cumulative distributions (symbols) for SRAM-sized transistors measured at three different readout times with the asymptotic approximation of Skellam distribution for large N (lines).

Besides, accordingly to eqn. 5, the ϕ value of 2 implies that $\mu_D=0$, i.e. there is no “Destruction” process for interface traps. The absence of interface traps repassivation is inline with the NBTI composite model findings for which the observed NBTI recovery is related to hole detrapping phenomenon [4]. Nevertheless, recent study shows ϕ value of about 2.7 (reported for reference in figure 9), which translates into a small but non negligible μ_D of about 0.12 to 0.20 [12]. In fact, there is no discrepancy here. Figure 13 shows comparison of cumulative distribution functions for non-relaxed (DC case) and relaxed (i.e. made up of only the permanent part) V_{TP} shifts with similar mean on identical devices. For non-relaxed devices, both permanent and recoverable parts play a role, as demonstrated by $\mu_D=0.15$ (i.e. $\phi=2.7$ inline with recent findings from [12]). For relaxed devices, degradation is dominated by interface traps which were permanently created during the stress, as demonstrated by $\mu_D=0$.

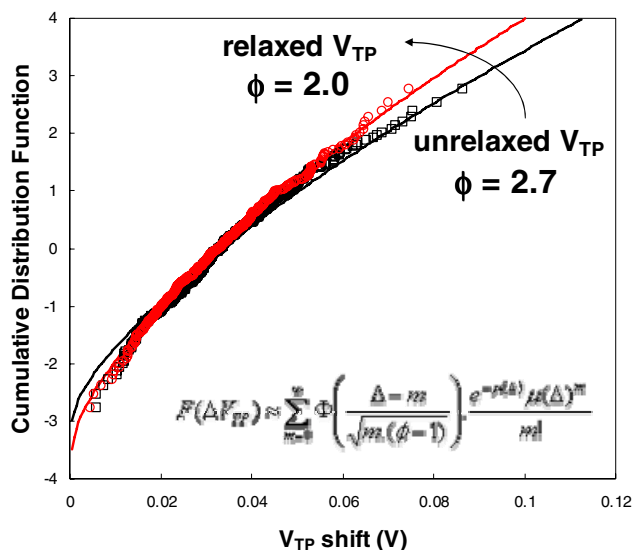
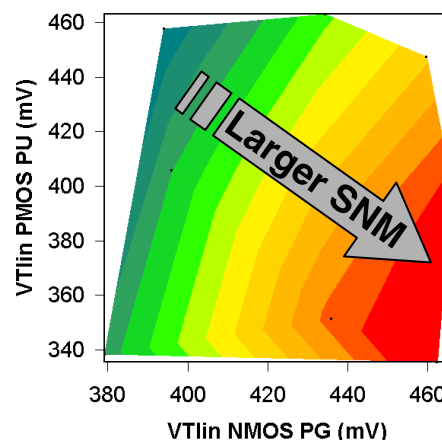


Figure 14: 6 transistor Static Random Access Memory (SRAM) bitcell schematic. Shaded transistor suffers NBTI degradation due to long-term storing of ‘0’ at BLTi node and ‘1’ at BLFi node.

Static Noise Margin (SNM) shows how stable the bitcell is during a reading operation. A large SNM corresponds to a stable bitcell.

Read current (I_{cen}) is the current discharged through the Bit-line (BL) when the cell is accessed. This parameter determines how fast the bitcell can be read, i.e. how performant is the bitcell. Therefore, a larger current is desirable.



V_{TP} shifts non-normality and SRAM

NBTI DEGRADATION: SRAM BITCELL STABILITY

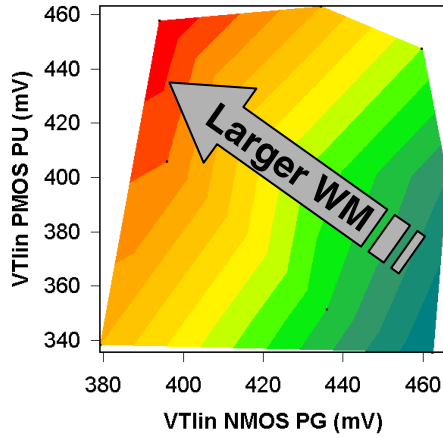


Figure 15: Impact of relative strengths of PD nMOS versus PU pMOS on SNM (top) and WM (bottom) as measured on SRAM bitcell using our 45nm process flow (Red color refers to larger values)

In this case, an increase of PU pMOS V_{TP} will yield to a modification of the relative strengths between nMOS and pMOS, thus resulting in a varied stability than originally designed. This modification can ultimately result in a bitcell failure, either during READ or WRITE operations.

READ operation

During READ operation, both Word-Line (WL) and BL are raised to “1”. Since access transistors PG are activated, the bitcell starts discharging to the BL, resulting in a voltage difference between the two BLS, further sensed by sense amplifier (Figure 16). A successful READ operation occurs when the bitcell discharging does not force the cell to flip, i.e. data loss. In parallel to the lowering of BL voltage, the inner nodes of the bitcell tend to be modified as well. The node storing “0” actually sees its potential increasing during READ operation. Since this node also controls the gate of the pMOS of the cross-coupled inverter, it might force this pMOS to switch of off-state, which will flip the cell. This cell flip will be eased if the pMOS has its V_{TP} increased due to NBTI.

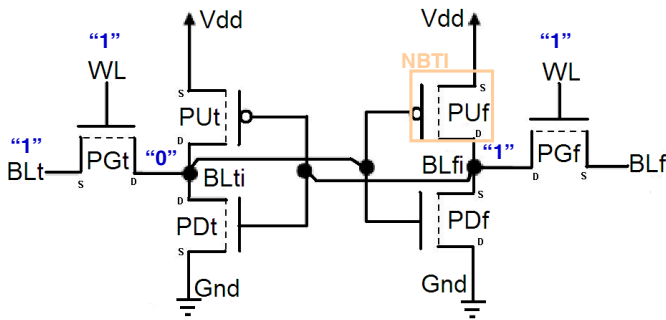


Figure 16: Long-term data retention (“0” in this schematic) followed by READ operation is the worst-case scenario for NBTI degradation. Shaded pMOS transistor is NBTI degraded.

A good metric to monitor the READ stability of a bitcell is the Static Noise Margin (SNM) of the cross-coupled inverters. SNM is generally defined as the largest fitted square in the wings of the transfer characteristics of both inverters. A READ failure is defined when the SNM value goes to zero.

From electrical analysis of a bitcell, it is expected that SNM value decreases when V_{TP} pMOS is increased. It has been observed both experimentally and through simulations using our in-house Design-in Reliability tools [13-14] that SNM shift linearly with respect to V_{TP} shift (Figure 17), inline with previous findings [15].

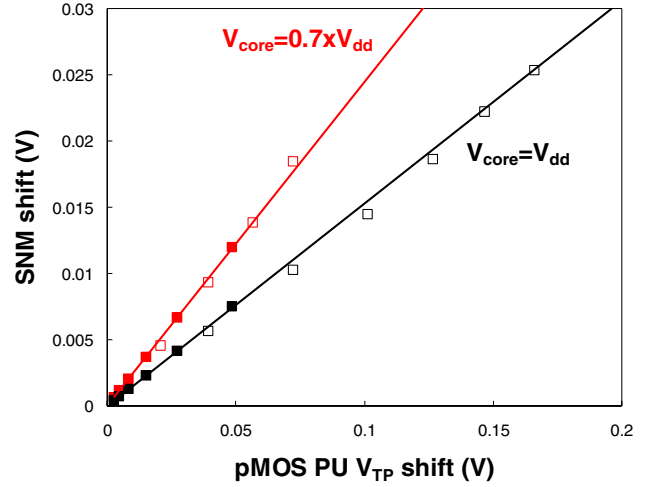


Figure 17: SNM increases linearly with pMOS PU V_{TP} shift, as experimentally measured on 65nm SRAM bitcell (Open symbols) and simulated using our Design-In Reliability tools (Filled symbols and lines). SNM sensitivity to similar V_{TP} shift is greater when V_{core} decreases.

The sensitivity of SRAM bitcell during READ operation to NBTI degradation can thus be written as:

$$\Delta SNM = -\frac{\partial SNM}{\partial V_{TP}}(V_{core}) \cdot \Delta V_{TP} = -dSNM \cdot \Delta V_{TP} \quad \text{eqn. 7}$$

where the correlation parameter dSNM (positive value) is dependent on the core voltage V_{core} supplying the array. This dependence needs to be modeled to further analyze the NBTI degradation of real product SRAM arrays.

WRITE operation

Writability of a bitcell is monitored using Write Margin (WM) metrics. Figure 18 summarizes the WRITE operation. During this operation, WL and BLf are raised to “1” and BLt is pulled down to “0”. The voltage for which the cell flips is defined as the Write Margin (WM). In real products, only a short period of time is allowed for WRITE operation. Ultimately, a WM value of zero means that it will take forever to write the bitcell. More generally, and depending on each product specifications and/or design, it is possible to define a realistic lower bound for WM for bitcell failure, which translates the allowed time window.

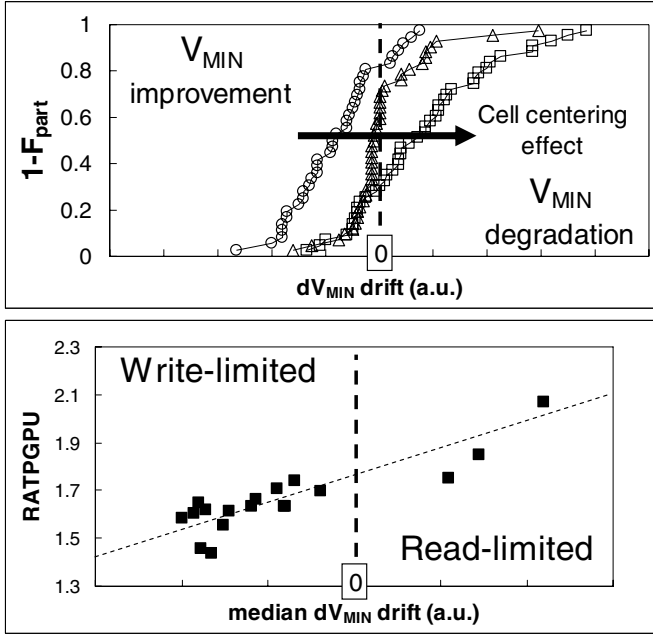


Figure 21: dV_{MIN} drift distributions for various cell centering (top). RATPGPU impact on mean dV_{MIN} drift for various lots (bottom).

In a product, there are several reasons, including extrinsic fails to observe a V_{MIN} drift. At this point, it is important to check if the observed V_{MIN} drift is linked to NBTI degradation. Figure 22 shows that for three different stress voltages, V_{TP} shift measured at device level and V_{MIN} drift monitored on real SRAM arrays during six months of test present the same time dynamics and the same voltage dependence. This result provides first insights that the mechanism responsible for the commonly observed V_{MIN} drift is the NBTI degradation.

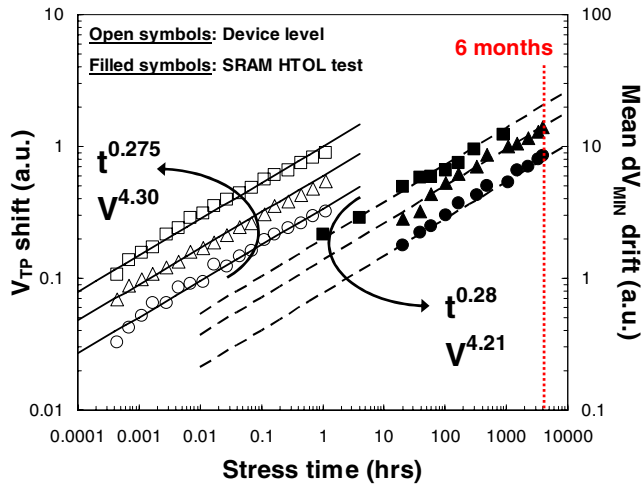


Figure 22: NBTI-induced V_{TP} shifts measured at device level (open symbols) and mean V_{MIN} drift at product level (filled symbols) for three identical stress voltages up to 6 months stress time. Identical time and voltage dependence are monitored.

Bitcell failure probability F_{bit}

A qualitative link between V_{MIN} drift and NBTI has been examined. Nevertheless, quantitative approach is needed to check with high accuracy if six-sigma confidence requirement is fulfilled and to allow the development of predictive tools. We propose here an approach based on the number of faulty cells present in a memory

cut. This approach has two main advantages: (1) to open the way to a direct link between bitcell/device and SRAM arrays levels and (2) to make quantitative predictions at the product end-of-life whatever the memory cut considered. This set of advantages can be further used to check if six-sigma confidence requirements are fulfilled and/or support process development to have realistic reliability guardband. The core element of this approach is the evaluation of the number of faulty cells in a memory cut at a given supply voltage. This number corresponds to a bitcell failure probability (F_{bit}).

Since NBTI does not degrade the read current I_{cell} , the modification of the number of bitcell failure can be related to either a read or a write failure. The method to determine the read failure probability will be thoroughly described here. The same approach will be followed to obtain the write failure probability.

As mentioned previously, the bitcell read failure probability $F_{\text{bit_read}}$ is directly related to SNM values. SNM is subject to random fluctuations in a similar way than device parameters such as threshold voltage V_{T} . Figure 23 shows SNM distributions for 45nm SRAM bitcell for various supply voltages, in good agreement with normal distribution. When SNM value is equal to zero, a read failure occurs when the bitcell is addressed as mentioned early in this work. This criterion allows determining the probability for which the read failure occurs at a given supply voltage.

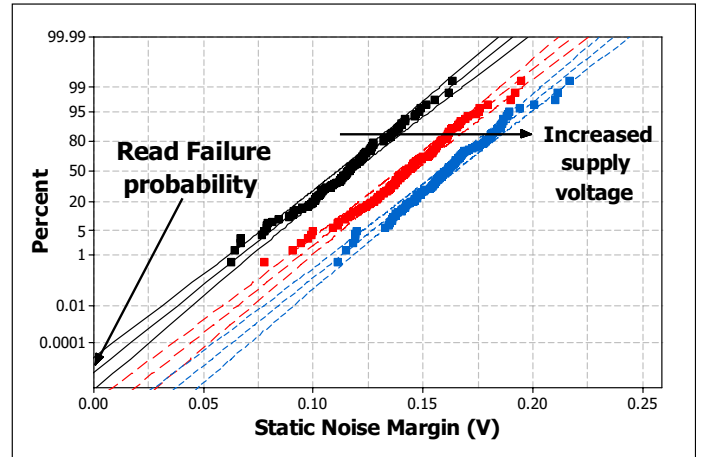


Figure 23: Static Noise Margin distributions for three different supply voltages showing normal distributions. Read failure probability is extracted when SNM value equals defined criterion, in this case zero value.

In a similar way, it is possible to define a criterion for the Write Margin WM which translates the operational time window allowed for writing. If the WM is below this criterion, the bitcell will take longer than the operational time window to be written correctly, i.e. a write failure occurs. This criterion is more design/operating test dependent than the SNM criterion and must be simulated.

Figure 24 shows the Read ($F_{\text{bit_read}}$) and Write ($F_{\text{bit_write}}$) failure probabilities simulated from isolated bitcells measurements in our 45nm process. Since a bitcell has either a read or a write failure, the two failure probabilities are independent and can be straightforwardly added. The resulting F_{bit} is the expected total bitcell failure probability as a function of the supply voltage.

In parallel, the number of faulty cells was monitored on SRAM arrays made of similar bitcells with the supply voltage sweeping from nominal voltage to defined lower bound. Knowing the total number of bitcells in the array, the F_{bit} is experimentally deduced at every single supply voltage value from the experimental number of

faulty cells on real arrays. Figure 24 shows excellent agreement with simulated F_{bit} for various centering corners.

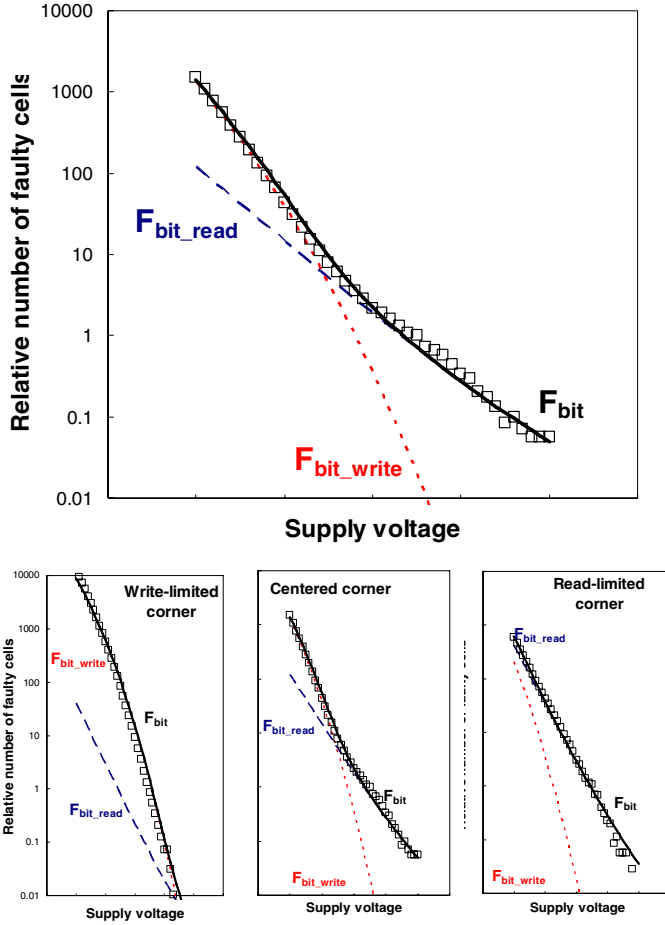


Figure 24: Top: Relative number of faulty cells measured experimentally in real SRAM arrays (open symbols) is in agreement with the predictions based on isolated bitcell measurements, according to the procedure described in text. Bottom: This approach has been validated on several corner splits.

Bitcell failure probability F_{bit} and NBTI degradation

The predictive simulations of the number of faulty cells in SRAM arrays show good agreement with non-degraded bitcell, i.e. at yield level. These simulations, as shown above, only require the knowledge of SNM and WM distributions. In a previous section of this paper, the dependence of SNM and WM to NBTI degradation was thoroughly examined. By combining these two approaches, simulations can be run to predict the evolution of the number of faulty cells during the operating life of SRAM arrays. Figure 25 shows an example of the output of these simulations. Non-degraded (i.e. fresh) bitcell failure probability is taken from previous centered corner (cf. Figure 24). We first simulated the impact of mean V_{TP} shift degradation, i.e. considering that NBTI does not add additional variations to the final V_{TP} distributions. Composite model was used to simulate accurately the mean V_{TP} shift value as a function of transistor/bitcell activity. In a second step, the increased dispersion induced by NBTI degradation is introduced, considering normal distribution at this point.

In the read-limited part (i.e. high supply voltage part), the mean V_{TP} shift increases the number of faulty cells, with respect to

degraded SNM. Additional introduction of NBTI-induced dispersion further increase the number of faulty cells.

In the write-limited part, the mean V_{TP} shift introduces a decrease of the number of faulty cells, which is linked to the improvement of WM. Nevertheless, this decrease is further compensated by the introduction of the NBTI-induced dispersion.

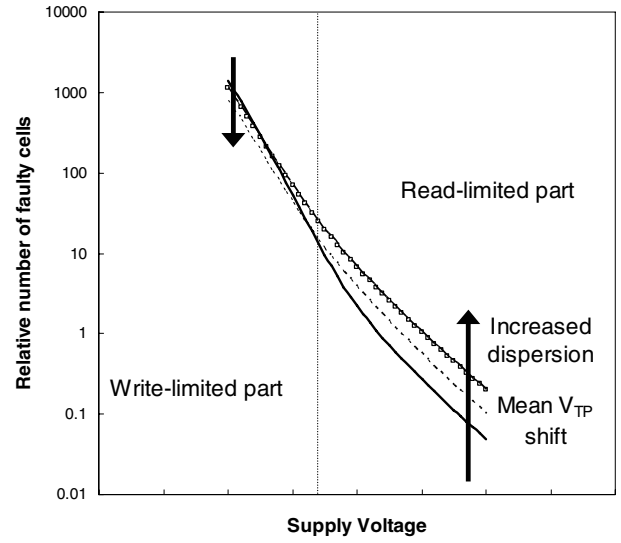


Figure 25: Simulated relative number of faulty cells for fresh bitcell (Full line), for NBTI-degraded bitcell without dispersion modification (dotted line) and NBTI-degraded bitcell with increased dispersion (small symbols).

Figure 26 shows that the simulated evolution of the number of faulty cells is in good agreement with experimental results from large SRAM arrays. To achieve these results, both the mean V_{TP} shift and the NBTI-induced dispersion have been modeled, as explained in figure 25. The good agreement between simulations and experimental results indicates that assuming normal distributions is a reasonable assumption considering our experimental accuracy.

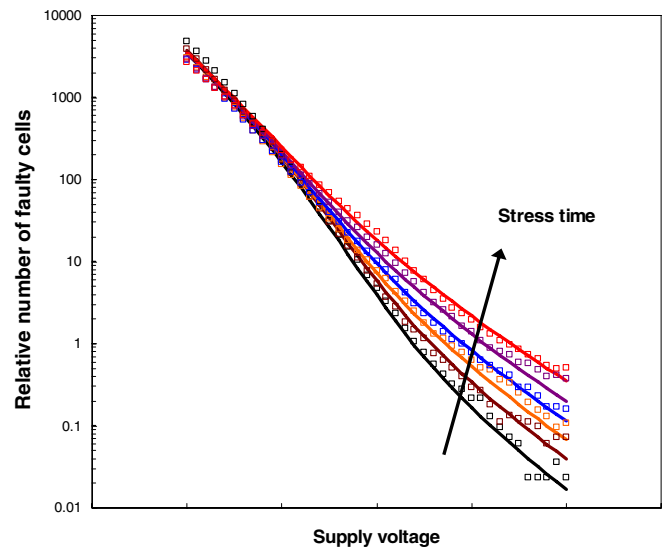


Figure 26: Simulations of the relative number of faulty cells (lines) compare accurately with experimental measurements made on large SRAM (symbols) for various readout times up to hundreds of hours.

From Bitcell failure probability F_{bit} to product V_{MIN}

So far, the number of faulty cells F_{bit} experimentally observed in SRAM arrays before and after operating life tests has been successfully modeled for the first time using predictive simulations based on elementary device/bitcell dataset. Nevertheless, product failure probability is of greatest interest for final customers. The probability F_{part} that SRAM array exceeds a requested V_{MIN} is:

$$F_{part} = 1 - (1 - F_{bit})^{N_{bits}} \quad \text{eqn. 9}$$

The combination of process choice and bitcell design determines a given F_{bit} at the requested V_{MIN} for the product mission profile. The number of bits allowed in the product presents an upper bound resulting from the combination of eqn. 9 and the six-sigma requirements (Figure 27).

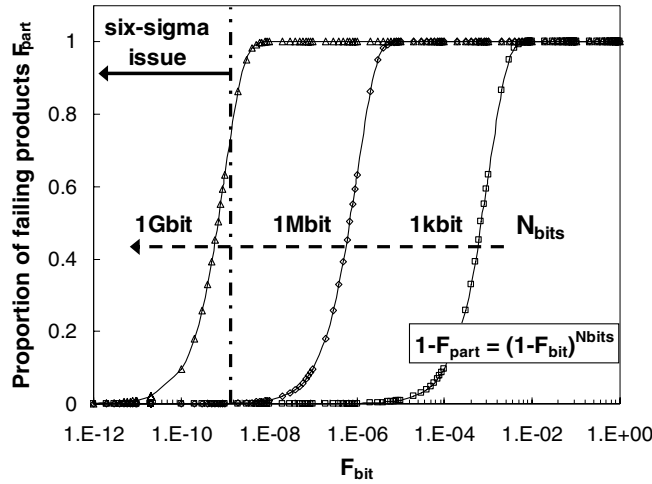


Figure 27: Proportion of failing products F_{part} as a function of the relative number of faulty cells F_{bit} , as a function of various cut sizes N_{bits} . Six-sigma confidence area is also shown for information.

The existence of the link between the number of faulty cells and the product V_{MIN} distributions gives some flexibility to support process development and/or design choice. Actually, it is possible to get accurate V_{MIN} distributions using only few parts but fully characterized through failed bits counting procedure. Figure 28 shows this direct link as well as a potential application to the case of SRAM bitcell transistors resizing.

Product screening and NBTI degradation

Despite all efforts to optimize SRAM bitcells to reach both yield and reliability within defined specifications, the ongoing scaling trend in critical dimensions and supply voltage will keep shrinking down the optimization window where both yield and reliability specifications are reached. In this context, screening of faulty products might be seen as a solution. Faulty products can be screened out either after initial readout (triangles in figure 29) or after short Burn-in (circles). Despite the initial improvement, screening procedure reveals to be inefficient to improve end-of-life V_{MIN} distributions, as seen in figure 29. Screening procedure artificially reduces the initial inherent variability. End-of-life V_{MIN} distributions are dominated by NBTI-induced variations, thus progressively masking all screening benefits until only NBTI dispersion dominates.

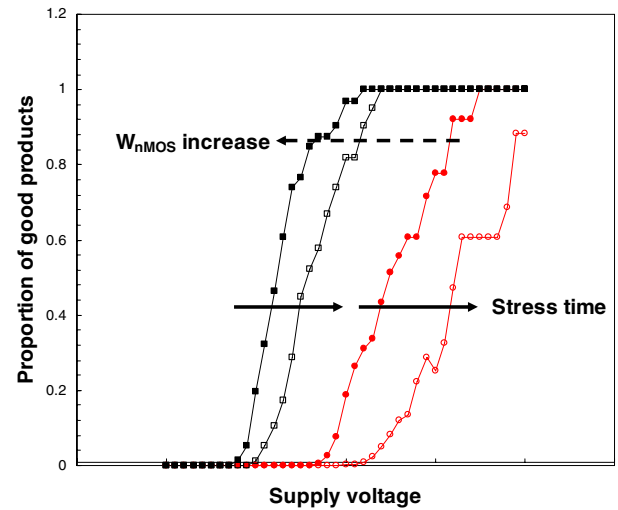
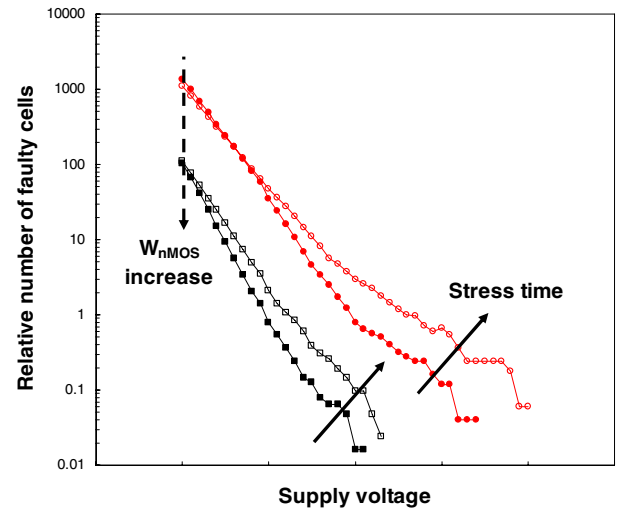


Figure 28: Top: Increased W_{nMOS} reduces the initial dispersion which reduces the number of faulty cells for fresh and degraded SRAM arrays. Bottom: This impact is also visible at V_{MIN} level.

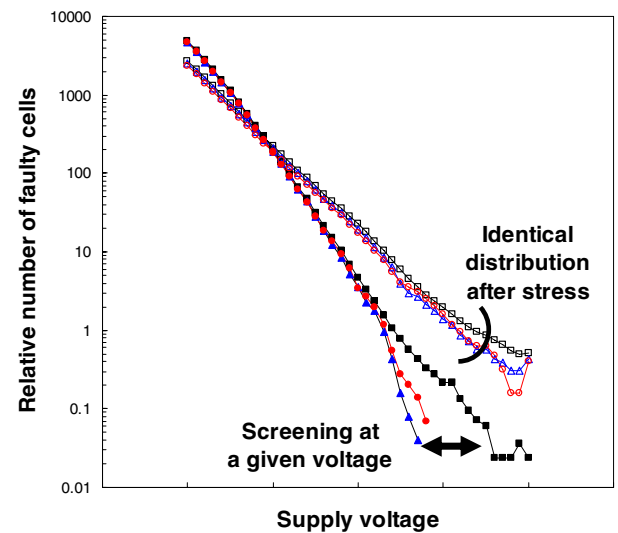


Figure 29: Impact of initial (triangles) and after Burn-In (circles) screening at a given voltage on fresh and degraded distribution of faulty cells. After degradation, no differences can be noticed.

Error Correcting code (ECC)

Nevertheless, some applications might require large memory capability (i.e. large number of bits), low failure rates and/or low end-of-life V_{MIN} . For such applications, Error Correcting Code (ECC) represents a satisfactory solution. ECC can correct product failures if they consist of single bits errors (SBE), such as in NBTI degradation case. Besides, the correction capability is quite high since with ECC-implemented a part fails only if 2 SBE occur in the same word of N -bits (N typically ranging from 64 to 256 bits). In the case of ECC-implemented, it is possible to redefine the product failure probability as:

$$F_{\text{part_ECC}} = 1 - (1 - F_{\text{word}})^{\frac{N_{\text{bits}}}{N}} \quad \text{eqn. 10}$$

which can be rewritten as [16]:

$$\begin{aligned} F_{\text{part_ECC}} &= 1 - (1 - F_{\text{bit}})^{N_{\text{bits}}} \left(1 + \frac{N \cdot F_{\text{bit}}}{1 - F_{\text{bit}}} \right)^{\frac{N_{\text{bits}}}{N}} \\ &\approx \frac{N-1}{2N_{\text{bits}}} (F_{\text{part_noECC}})^2 \end{aligned} \quad \text{eqn. 11}$$

ECC allows important reduction of the proportion of failing products, roughly equivalent in our example to hundred times reduction in cut size. Of course, in real products, part of the ECC capability is used to reduce the impact of extrinsic issues as well as to achieve acceptable yield. Nevertheless, since the correction power of ECC is quite large, it is reasonable to think that it might help to solve out some NBTI issues in specific, reliability-oriented applications.

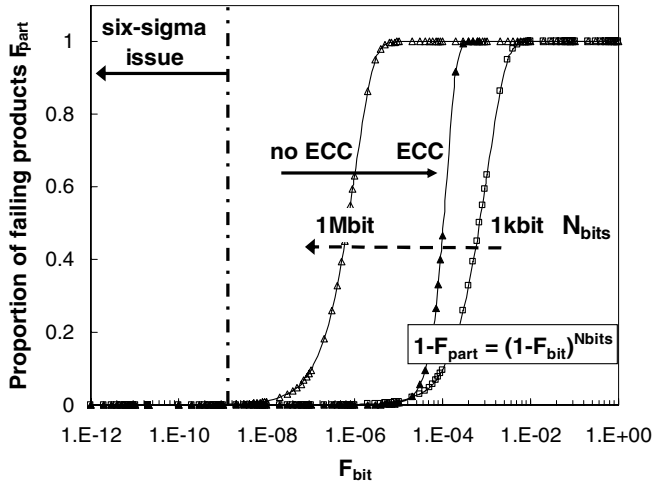


Figure 30: Proportion of failing products F_{part} as a function of the relative number of faulty cells F_{bit} for two SRAM cut sizes: 1kbit (squares) and 1Mbit (triangles). Improvement of using ECC (filled symbols) versus without ECC (open symbols) (ECC: 128 bits/word) is shown for 1Mbit array.

SRAM access time

In high-performance SRAM products, the V_{MIN} is not the only parameter of importance. Not only the SRAM array should maintain the data during read/write operations, but the data should be accessed

with high speed. The time delay from the clock edge to valid data presented on the bus is defined as access time (T_{aa}). In addition to V_{MIN} drift, NBTI degradation will also slow down this time delay. We have simulated the degradation of this access time in real SRAM products as a function of mean V_{TP} shift. In parallel, the access time degradation was monitored experimentally for three various stress voltages on simulated products. Figure 31 shows the accuracy of the simulations versus the experimental dataset. Projections of the access time degradation under operating conditions showed that the degradation remains confined within the design guardband up to specified product field lifetime.

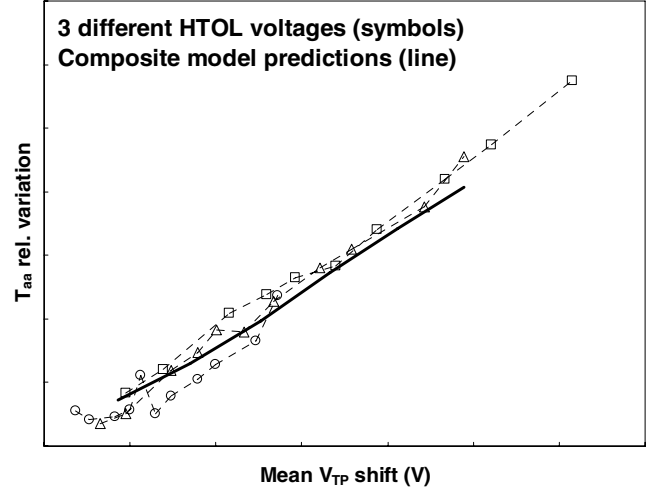


Figure 31: Simulated access time T_{aa} relative degradation as a function of increased mean V_{TP} shift (lines) on real 65nm product. Experimental results (symbols) are shown for comparison for three various stress voltages.

CONCLUSIONS

A novel composite model had been recently introduced to physically explain the mean pMOS threshold voltage shift (V_{TP}) induced by NBTI degradation at transistor level in a quantitative way. In this study, the statistical aspect of NBTI degradation has been developed in the framework of NBTI composite model. Through a thorough study of isolated SRAM bitcell sensitivity to NBTI degradation, NBTI degradation of large SRAM arrays have been accurately modeled. Comparison with large experimental datasets shows excellent agreement with simulations. This approach allows quantitative prediction of NBTI-induced V_{MIN} variations and access time T_{aa} degradation during burn-in operations. The key findings include (a) demonstration of non-normality of V_{TP} shift distribution (b) NBTI contribution to product V_{MIN} drift arises from both mean V_{TP} drift but also from increased V_{TP} dispersion, and (c) V_{TP} shift non-normality is smoothed out at product level by time-zero variation of the six transistors of the SRAM bitcell.

REFERENCES

- [1] R. Kapre *et al.*, "SRAM variability and supply voltage scaling challenges", in *Proceedings of IRPS*, 2007, p. 23.
- [2] S. Mukhopadhyay *et al.*, "SRAMs in scaled technologies under process variations: failure mechanisms, test & variation tolerant design", in *Proceedings of CICC*, 2006, p. 547.
- [3] S. Mahapatra, "NBTI in pMOSFETs: characterization, modeling and material dependence", in *Proceedings of IRPS*, 2007, tutorial 111 and references therein.
- [4] V. Huard *et al.*, "New characterization and modeling approach for NBTI degradation from transistor to product level", in *Proceedings of IEDM*, 2007, session 31
- [5] B. Kaczer *et al.*, "A disorder-controlled-kinetics model for Negative Bias Temperature Instability and its experimental verification", in *Proceedings of IRPS*, 2005.
- [6] V. Huard *et al.*, "NBTI degradation: From physical mechanisms to modeling", in *Microelectronics Reliability*, 2006.
- [7] T. Grassler *et al.*, "The universality of NBTI relaxation and its implications for modeling and characterization", in *Proceedings of IRPS*, 2007.
- [8] C. Schluender *et al.*, "Consideration of recovery effect during NBTI measurements for accurate lifetime predictions of state-of-the-art pMOSFETs", in *Proceedings of IRPS*, 2007.
- [9] V. Huard *et al.*, "Single hole detrapping events in pMOSFETs NBTI degradation", in *Proceedings of IRPS*, 2005.
- [10] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μ m MOSFETs: a 3-D atomistic simulation study", in *Transactions on electron devices*, 1998, p. 2505.
- [11] J. G. Skellam, "The frequency distribution of the difference between two Poisson variates belonging to different populations", in *Journal of royal statistical society, A*, vol 109, 1946, p. 296.
- [12] S. Rauch, "Review and reexamination of reliability effects related to NBTI-induced statistical variations", in *Transactions on device and materials reliability*, 2007, p. 524.
- [13] C. Parthasarathy, "Design-In Reliability (DiR) with emphasis on NBTI", in *Proceedings of IIRW*, tutorial, 2005.
- [14] V. Huard *et al.*, "Design-in Reliability approach for NBTI and Hot-Carrier degradations in advanced nodes", in *Transactions on device and materials reliability*, 2007, p. 558.
- [15] A.T. Krishnan *et al.*, "SRAM cell static noise margin and V_{MIN} sensitivity to transistor degradation", in *Proceedings of IEDM*, 2006, session 12.
- [16] A. Haggag *et al.*, "Understanding SRAM High-Temperature-Operating-Life NBTI: Statistics and permanent vs recoverable damage", in *Proceedings of IRPS*, 2007.