HTOL SRAM Vmin shift considerations in scaled HKMG technologies

S. Balasubramanian, V. Joshi, T. Klick*, R. Mann#, J. Versaggi#, A. Gautam#, C. Weintraub+
G. Kurz*, G. Krause*, A. Kerber#, B. Parameshwaran, T. Nigam
GLOBALFOUNDRIES, Santa Clara CA, *Dresden, Germany, #Malta, NY, +Austin, TX

***Abstract*—This paper examines the role of NBTI and PBTI on SRAM Vmin shifts during HTOL stressing and quantifies their impact on reliability lifetime projections in scaled high-k metal gate (HKMG) technologies. Correlation between measured HTOL SRAM Vmin shifts and transistor level parametrics is summarized on both 28nm poly-SiON and HKMG technologies. The paper concludes that the commonly used HTOL acceleration voltage of 1.4xVnom may be excessive in scaled HKMG technologies due to the larger role of PBTI in SRAMs**

***Index Terms* - SRAM, Vmin, HTOL, NBTI, PBTI, HKMG, replacement metal gate, scaling, CMOS**

I. Introduction

In aggressively scaled technologies, increased threshold voltage (Vt) variations and device reliability wear out mechanisms present significant challenges to SRAM Vmin scaling. Given SRAM devices are scaled significantly in every node for density reasons, random dopant fluctuations (RDF) is the dominant mechanism limiting threshold voltage (V_T) variations in these devices. This in turn limits SRAM Vmin scaling. Since the 130nm technology node, SRAM Vmin scaling is not only limited by time zero Vt variations (RDF) but also by time-dependent Vt increase in median and sigma due to device wear out. In the poly SiON technologies, the key limiter for Vmin shifts in SRAMs is the NBTI degradation in PU devices. With the introduction of HKMG, PBTI has emerged as an added reliability wear out mechanism in addition to NBTI. PBTI and NBTI manifest themselves as both systematic Vt shifts (median) as well as increased BTI related statistical Vt variations [1,2]. Therefore, in order to optimize SRAM end-of- life (EOL) Vmin, it is necessary to consider both time-zero (T_0) yield as well as NBTI/PBTI induced Vmin degradation. High Temperature Operating Life (HTOL) stress testing is used in product and technology qualification to quantify EOL Vmin shifts. This paper examines the continued applicability of the historic standard HTOL stress conditions (1.4xVnom/ 1000h/T≥125C) to scaled HKMG SRAMs.

The paper is organized in 4 sections. First we illustrate the impact BTI on SRAM cell operation in terms of Vmin shift. We discuss the qualitative role played by NBTI and PBTI in the degradation of SRAM cell. This is followed by discussion of the simulation methodology used to correlate to Si-measured EOL Vmin degradation as well as predict the same for scaled SRAMs. The methodology is calibrated to both Poly SiON technology and HKMG based technology in 28nm. The impact of PBTI vs NBTI on a given HTOL Vmin shift is further discussed in the context of voltage acceleration coefficient differences for the two mechanisms. Finally guidance is provided for a technology that just meets both the NBTI and PBTI specs. It is shown that in such cases the historical approach for Vmin guard band is overly conservative and unrealistic. We conclude by proposing a stressing methodology for appropriate parametric Vmin guard banding thereby validating EOL product functionality during HTOL.

II. Role of BTI on SRAM Read/Write Margin Shifts

A. Impact of BTI on Vmin shift in SRAMs

Vmin is defined as the lowest voltage where the SRAM array is functional with no fails during write and read back of the whole array. During HTOL, packaged SRAM's are stressed dynamically by writing/reading of different patterns to the SRAM array with Vmin characterized pre- and post-stress. In addition - SRAM read disturb voltage (VdipR) and write margin (WMar) are also monitored as a function of stress time.

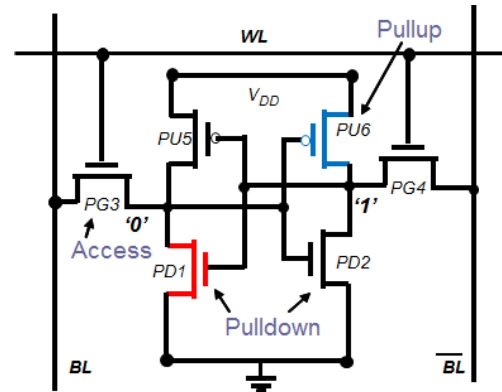

Fig. 1. Schematic of a 6-T SRAM cell. When the cell stores the state '0', the transistors PD1 (pull down NMOS) and PU6 (pull-up PMOS) experience PBTI and NBTI stress respectively. When the state toggles to '1', then the transistors PD2 and PU5 experience BTI stress in turn.

	Poly/SiON (NBTI only)	HKMG (PBTI+NBTI)
VdipR (Read disturb)	Degrades- due to PU NBTI	Degrades – due to a) PU NBTI b) beta ratio (PD/PG) degrades c) PD Vtmm increase
WMar (Write Margin)	Improves - due to PU NBTI	Improves/Neutral - due to PG/PU V_t increase.

Table 1: Impact of NBTI and PBTI on SRAM read disturb and Write Margins. VdipR degrades whereas WMar improves or stays neutral with HTOL stress. The VdipR degradation with HTOL stress is the primary reliability related concern for SRAM functionality.

978-1-4799-3286-3/14/$31.00 ©2014 IEEE

The overall Vmin of the SRAM array is the maximum of read vs write Vmins. The theoretical minimum Vmin occurs when read and write Vmin values are balanced. However, with BTI induced degradation, the read vs write Vmin balance changes through the lifetime of the product.

Fig. 1 shows an SRAM cell storing the data state '0'. Under this condition, transistors PD1 (pull down NMOS) and PU6 (pull-up PMOS) experience PBTI and NBTI stress respectively. When the state toggles to '1', then PD2 and PU5 experience BTI stress in turn. Table 1 qualitatively summarizes the impact of NBTI and PBTI shift mechanisms on SRAMs. NBTI shifts result in a VdipR degradation and WMar improvement. In HKMG SRAMs, in addition to VdipR degradation from NBTI, PBTI introduces additional VdipR degradation due to a reduction in SRAM cell beta ratio (Idsat,PD/Idsat,PG) and PD Vt mismatch (Vtmm) degradation. The PG device only experiences PBTI stress during a write '0' cycle and therefore tends to degrade much less than the PD device. This results in a beta ratio reduction, thereby causing degradation in VdipR.

B. Vmin distribution pre and post stress

Fig. 2 shows an example 28nm HTOL Vmin chart from a 64Mb array stressed at 1.4xVnom/125C. Vmin and VdipR distributions are plotted for both T0 and after 1000hr stress (T1000). VdipR degrades during HTOL stressing due to PU Vt increase in line with the expectations summarized in Table 1. On the other hand, the overall Vmin distribution shifts towards lower values with HTOL stress. The reason for this is that in this particular case, the overall Vmin of is limited by SRAM write margin, WMar, at both T0 and post stress. Given the PU NBTI degradation with stress, the Vmin improves with HTOL stress due to improvements in WMar. The impact of higher VdipR is negated by the improvement in WMar at 85C. In effect, this cell comfortably met the time zero and EOL Vmin spec. The read vs write Vmin balance is determined by a number of technology specific factors such as performance requirements, leakage, array organization (column height), planned product usage and acceptable production Vmin screens.

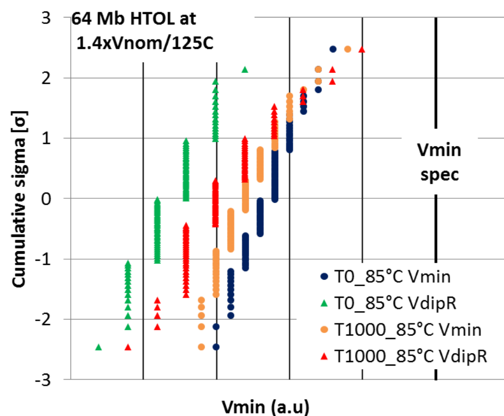

Fig. 2. A representative chart of 28nm HTOL Vmin shifts showing that VdipR degrades with HTOL stress. In this case, the overall Vmin is limited by write margin, WMar, at both T0 and post stress, which in turn improves with HTOL stress.

III. SIMULATION METHODOLOGY AND CALIBRATION

A. HTOL Vmin correlation to NBTI/PBTI

This section establishes correlation between measured HTOL parametric Vmin shifts (ΔVmin) and that simulated using measured transistor level NBTI /PBTI median and sigma ΔVt shifts. The Vmin simulations consider both the T_0 transistor targeting (Vt, transistor performance) as well as variations such as Vt, Lg etc. Time zero (T0) Vmin simulations are performed across various temperature and voltages to calibrate with the measured T0 Vmin/VdipR distributions. The impact of BTI-induced Vt shifts is then incorporated in the simulations, where both the increase in median Vt and any impact of additional variability due to BTI is included. The BTI Vt degradation inputs for median and sigma include the impact of BTI recovery as well to capture the effect of data toggling during functional array stress. A schematic flow of the simulation methodology is shown in Fig. 3

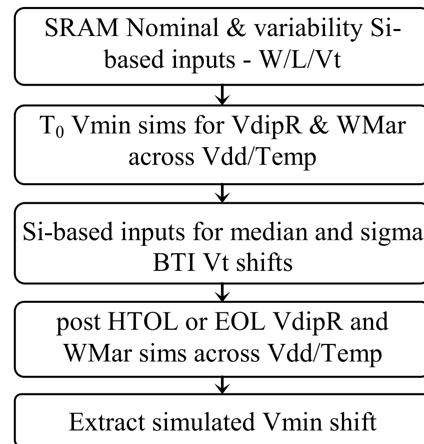

Fig. 3. Steps in the HTOL Vmin shift simulation/ correlation methodology

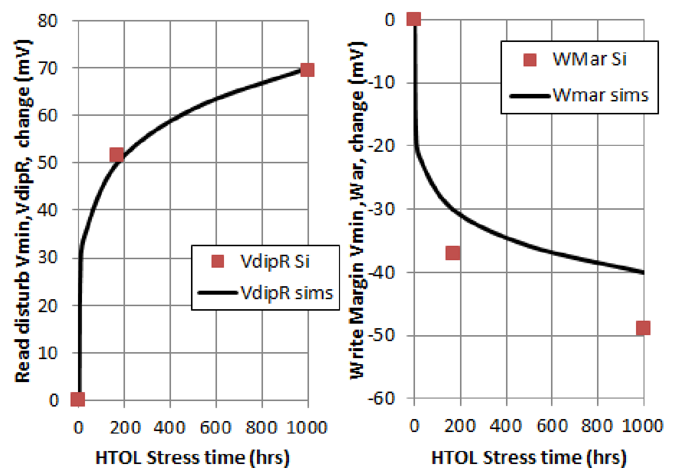

Fig. 4. 6T SRAM a) read disturb Vmin distribution (VdipR) at 125C from 28nm HKMG technology (TechA) showing progressive degradation in VdipR with increasing stress b) Write margin, WMar, measured at -40C shows improvement

B. Calibration to measured HTOL Vmin shift data

The ΔVmin correlation to gate-first 28nm HKMG (TechA) Si stressed at Vstress = 1.4xVnom, T=125C is shown in Fig. 4. The data and simulations both show that the VdipR at 125C (worst case read disturb temperature) increases with HTOL stress whereas the measured WMar at -40C (worst case test temperature for writability) decreases with stress. The simulations show good correlation to Si data for both read and write Vmin changes with HTOL stressing. Based on these correlations, we conclude that BTI plays a significant role in SRAM HTOL Vmin shifts and that the simulation framework captures the HTOL Vmin shifts quantitatively.

IV. HTOL VMIN SHIFT DATA AND SCALING TRENDS

A. HTOL Induced Vmin Shift For 28 nm Technologies

A common foundry accepted guard band for EOL Vmin shift is ΔVmin=5%Vnom. For example, Fig. 5 shows HTOL Vmin shifts from a 28nm poly-SiON technology which meets this criterion, with NBTI being the major HTOL Vmin limiter.

In another 28nm gate-first HKMG optimized for low-Vdd performance (TechB), we still comfortably meet the ΔVmin spec in spite of the additional degradation from PBTI. The simulations for Vmin and VdipR shift are well matched to the Si data and total EOL Vmin shift < 50mV (Fig. 6). This is because, for this HKMG technology, little PBTI is seen even under HTOL stressing of 1.4xVnom, so Vmin continues to be limited by NBTI. The low PBTI in gate-first HK is attributed to the presence of La for work function engineering [3]. For both these technologies ΔVmin was determined completely by ΔVdipR which shows degradation as a function of stress time as explained in Table I.

B. NBTI vs PBTI Tradeoff During HTOL Stress

Fig. 7 illustrates the challenge and critical role PBTI plays in HKMG technologies under HTOL stressing conditions. A higher voltage acceleration exponent (VAE) in PBTI as compared to NBTI [4,5] means that EOL equivalent BTI Vt shifts are generated at lower stress voltages in PBTI than for NBTI. This implies that the choice of stress voltages is a critical HTOL consideration in scaled HKMG technologies. At short HTOL stress times (<24hrs), there is a relatively large Vstress difference (0.2-0.3xVnom) to meet a certain BTI Vt shift target due to the NBTI vs PBTI VAE differences. At longer stress times (~1000hrs), the Vstress gap needed to induce the target NBTI vs PBTI Vt shifts is much smaller (~0.1xVnom). This implies lower Vtstress applied for longer stress times induces a better balanced PBTI vs NBTI Vt shifts, thereby better mimicking EOL behavior. However, in 28nm gate-first HKMG, under typical HTOL conditions of 1.4xVnom, the PBTI Vt shift << NBTI even after accounting

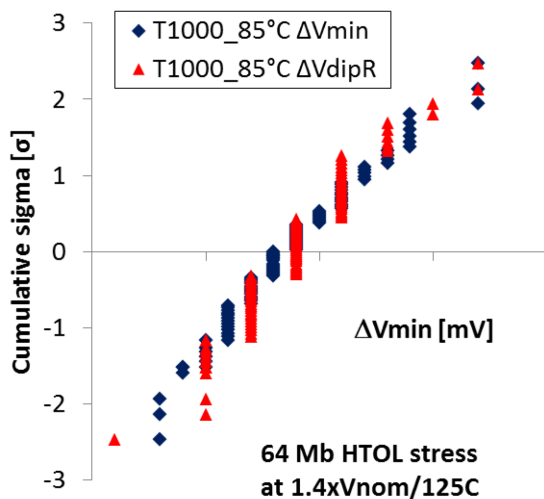

Fig. 5. 28nm poly-SiON HTOL Vmin shift data showing that the VdipR shifts are in the range of ~5%Vnom after stressing at 125C/ 1.4xVnom/1000h. The Vmin shifts are driven primarily due to PU NBTI degradation

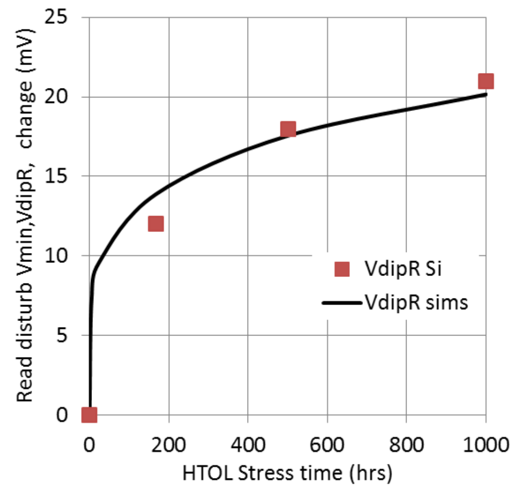

Fig. 6. 28nm gate-first HKMG (TechB) HTOL data showing VdipR shifts of <50mV after stressing at 1.4xVnom/1000h/ 125C. The Vmin shifts are still driven primarily due to NBTI degradation given gate-first HKMG showed little PBTI even at the HTOL stress voltage

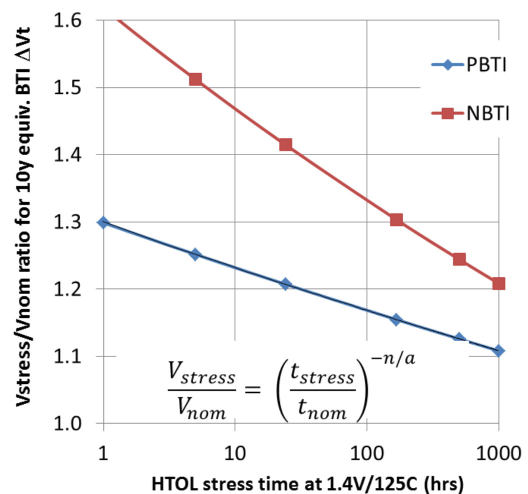

Fig. 7. Voltage acceleration ratio (Vstress/Vnom) required to mimic EOL-like Vt shifts. PBTI has a steeper voltage acceleration (VAE factor a) than NBTI and therefore needs a lower stress voltage to produce the same BTI Vt shift.

for differences in their VAE. This is just an illustration of how the traditional 1.4xVnom has not been an issue in 28nm because of good PBTI control.

V. HTOL VMIN IMPLICATION FOR HKMG

With continued Tinv scaling, controlling BTI to meet the device reliability specs at Vmax, 125C, 5yrs DC is a significant challenge [6]. We study a HKMG technology scaling scenario, where one just meets the NBTI and PBTI device specs for read-disturb Vmin limited cell. For reference, the 28nm poly-SiON & gate-first HKMG summarized above bettered this device BTI spec. Using the HTOL simulation methodology calibrated on 28nm poly-SiON and HKMG, our simulations project that a traditional HTOL stress of 1.4x Vnom would produce EOL-like Vmin shifts in only ~1h of stressing (Fig. 8). This is primarily driven by a higher PBTI under HTOL stress voltages arising from a steep VAE and larger sensitivity of SRAMs to PBTI. A more reasonable HTOL stress is depicted by the grey box in Fig. 8 (for example, 1.25xVnom/168h), which better captures EOL ΔVmin targets. We examine the choice of HTOL Vstress needed to meet EOL ΔVmin targets across BTI scaling scenarios, namely - NBTI limited vs. PBTI limited.

Fig. 9 shows that while HTOLVstress~1.4xVnom is acceptable for NBTI limited technologies, Vstress~1.25xVnom is more relevant for a case where NBTI and PBTI are roughly equal. Based on these projections, the continued use of HTOL stressing at 1.4xVnom makes meeting the ΔVmin criteria difficult unless the PBTI magnitude can be significantly reduced. However, our recommendation of Vstress~1.25x Vnom is only for intrinsic or parametric Vmin shifts and continuing HTOL at higher Vstress or longer times would be prudent for capturing extrinsic fails.

VI. CONCLUSION

Given the larger sensitivity of SRAMs to PBTI and SRAM scaling, the continued use of the traditional HTOL stress voltages ~1.4xVnom/1000h in scaled HKMG technologies will result in large SRAM Vmin shifts that are significantly greater than that expected from EOL usage. HTOL readouts at lower stress voltages and times (~1.25xVnom/168h) should be considered for appropriate parametric Vmin shift coverage from stressing and continued HTOL at higher Vstress or longer times would be appropriate for capturing extrinsic fail components.

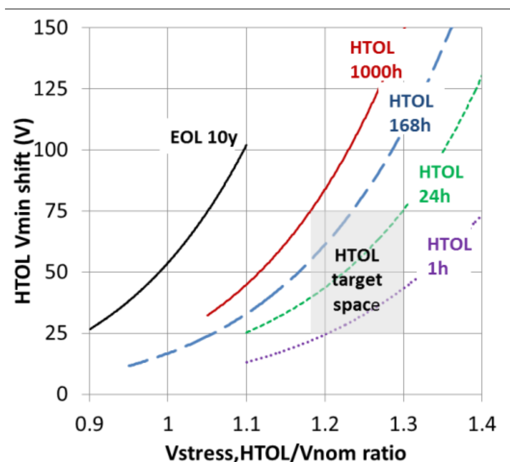

Fig. 8. Projections of SRAM Vmin shifts vs. HTOL stress voltage for a technology that just meets the transistor level BTI specs. At Vstress = 1.4xVnom, target EOL-like Vmin shifts are produced already in ~1h of stressing. A more reasonable HTOL stress condition is highlighted in the grey box with Vstress between 1.2x Vnom and 1.3xVnom.

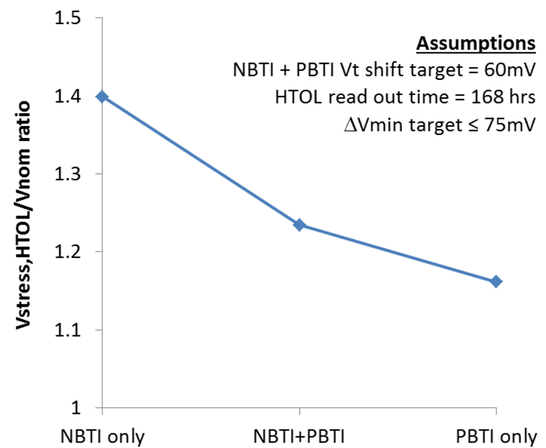

Fig. 9. Projected HTOL voltage stress (Vstress/Vnom ratio) to cover EOL ΔVmin targets across BTI scaling scenarios, namely - NBTI limited vs. PBTI limited technologies. In the NBTI limited case, HTOL at Vstress/Vnom~1.4x is acceptable, whereas for a PBTI only case, this ratio drops to ~1.2x. For a scenario where NBTI and PBTI are ~equal, then HTOL at Vstress/Vnom ~ 1.25x looks reasonable.

REFERENCES

[1] Kapre, R., K. Shakeri, H. Puchner, J. Tandigan, T. Nigam, K. Jang, M. V. R. Reddy, S. Lakshminarayanan, D. Sajoto, and M. Whately. "SRAM variability and supply voltage scaling challenges." IEEE IRPS, pp. 23-28., 2007

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

[3] Sato, M., N. Umezawa, J. Shimokawa, H. Arimura, S. Sugino, A. Tachibana, M. Nakamura et al. "Physical model of the PBTI and TDDB of la incorporated HfSiON gate dielectrics with pre-existing and stress-induced defects." IEEE IEDM, pp. 1-4., 2008.

[4] Pae, S., Choi, J., He, J., & Lee, S. H. (2010). "Reliability characterization of 32nm high-K and Metal-Gate logic transistor technology." IEEE IRPS, pp. 287-292. 2010.

[5] Rahman, A., M. Agostinelli, P. Bai, G. Curello, H. Deshpande, W. Hafez, C-H. Jan et al. "Reliability studies of a 32nm System-on-Chip (SoC) platform technology with 2 nd generation high-k/metal gate transistors." IEEE IRPS, pp. 5D-3., 2011.

[6] Cartier, E., A. Kerber, T. Ando, M. M. Frank, K. Choi, S. Krishnan, B. Linder et al. "Fundamental aspects of HfO 2-based high-k metal gate stack reliability and implications on t inv-scaling." IEEE IEDM, pp. 18-4. IEEE, 2011.