F_{MAX} / V_{MIN} and noise margin impacts of aging on domino read, static write, and retention of 8T 1R1W SRAM arrays in 22nm high-k/metal-gate tri-gate CMOS

Jaydeep P. Kulkarni, Carlos Tokunaga, Minki Cho, Muhammad M. Khellah, James W. Tschanz, and Vivek K. De
Circuit Research Lab, Intel Corporation, Hillsboro, OR, USA (E-mail: jaydeep.p.kulkarni@intel.com)

Abstract

Progressive impacts of aging on F_{MAX} & noise margin of the precharge-evaluate domino read, and V_{MIN} for differential static write & retention are demonstrated via statistical measurements over the operational lifetime of a 14KB 1R1W 8T SRAM array in 22nm high-k/metal-gate tri-gate CMOS.

Introduction

Voltage-temperature stress-induced aging due to BTI (Bias Temperature Instability) and HCI (Hot Carrier Injection) in scaled high-k/metal-gate CMOS impacts maximum frequency (F_{MAX}), minimum supply voltage (V_{MIN}) and noise margin of digital logic and SRAM arrays in SoC IP blocks over the operational lifetime. Typically, the worst-case aging impacts are included in the F_{MAX}/V_{MIN} settings as guardbands, and the circuits are designed for adequate noise immunity in the presence of worst-case aging degradations. Thus, significant power/performance overheads are incurred over the SoC lifetime. Various on-die aging monitors have been proposed to track aging impacts on the performance of static CMOS logic critical paths [1-3]. Overall aging impacts on F_{MAX}/V_{MIN} and robustness of an IP block are governed, however, not only by the static CMOS logic circuits, but also by the 8T SRAM arrays with single-ended full-swing domino read and differential static write operation. In this paper, we demonstrate by direct measurements and statistical analysis, progressive aging impacts on (1) F_{MAX} of the precharge-evaluate domino read, (2) read noise margin, and (3)V_{MIN} for differential static write (with assist) and retention, over the operational lifetime of a 14KB 1R1W 8T SRAM fabricated in 22nm high-k/metal-gate tri-gate CMOS (Fig. 7).

Array Stress Scenarios

Measurements are performed for both active and idle stress conditions to capture effects of time-dependent aging/recovery, and cover the entire range of stress scenarios encountered in normal operation. In idle mode, where the domino read path is in precharge state, local bitline (LBL) precharge PMOS & keepers, LBL merge NAND pulldown NMOS, global bitline (GBL) precharge PMOS, feedback inverter pulldown NMOS, and set dominant latch (SDL) pulldown NMOS are stressed (Fig. 1). During active read mode, read wordline (RWL) NMOS, LBL merge PMOS, GBL pulldown NMOS, feedback inverter pullup PMOS, and SDL pullup PMOS are stressed. For worst-case aging impact on read F_{MAX} (read noise margin), the read port bit NMOS N2 is stressed with a stored bitcell value of '1' ('0'). A stored '1' is used for worst-case aging impacts on static write and retention V_{MIN}. In idle mode, the bitcell PMOS PL holding '1', the bitcell NMOS NR holding '0', and the PMOS S1 are stressed. The always-on minimum size PMOS S1, connected between Vcc and the bitcell supply (BitVcc) and shared among multiple cells, is used as write assist [4]. During active write mode, the write port NMOS devices are also stressed.

Array Aging Measurements

Each 4Kb array of the 14KB 1R1W 8T SRAM macro is organized as 128 entries X 32 I/Os (Fig. 2, 7). The chip is stressed at 1.2V/90^{0}C for a duration equivalent to 1.5 years of lifetime. Read, write, and retention F_{MAX}/V_{MIN} measurements are performed at nominal voltages after each stress interval. This stress and measurement sequence is repeated six times to capture the progressive impacts of aging for a cumulative stress duration equivalent to 10.5 years of lifetime. The lower half of the array is characterized for idle mode stress, with R/WWL<63:0> held at '0' during the entire stress interval (Fig. 2). The upper half of the array is characterized for active mode read/write stress where R/WWL<127:64> are stressed sequentially, and this cycle is repeated for the entire stress interval. The stress duration is progressively higher along the domino read path. PMOS driving the keepers in the LBL keeper nodes, shared among the 16 bitcells on each LBL, are subjected to 16 times longer stress than the RWL NMOS. Similarly, transistors at the LBL merge NAND (SDLOUT) output nodes are stressed 32 (64) times longer.

Statistical Array Measurement Results

Aging behaviors of 40 randomly selected bits across the arrays are tracked for read, write & retention from the beginning-of-life (BOL) to end-of-life (EOL) to reveal impacts and interactions of static within-die (WID) device parameter variations at BOL and the statistical time-dependent aging/recovery mechanisms from BOL to EOL (Fig. 3). Read F_{MAX} improves while read noise margin degrades with aging, due to weakening of contentions from the keeper PMOS and other stressed transistors along the critical domino read evaluate path. Unlike static CMOS logic where the critical path delay typically worsens with AC aging stress, the critical evaluate delay in a domino precharge-evaluate path actually improves with aging at the expense of degraded noise margin. For some bits, write V_{MIN} improves due to reduced contention from the weakened bitcell PMOS/NMOS and the larger droop on BitVcc node induced by the weakened minimum-sized write-assist PMOS. For some other bits, write V_{MIN} remains approximately unchanged due to compensating effects of the weakened write-assist PMOS on write completion, especially for active mode stress where the write port NMOS devices in the cell also weaken, thus further compensating for reduced contention in the bitcell. Retention V_{MIN} remains the same or degrades with aging. No significant correlation between the amount of WID parameter variation impact at BOL and the degree of aging-induced change from BOL to EOL is observed for F_{MAX}/V_{MIN}, noise margin, or retention, thus confirming the uncertain & time-dependent nature of aging/recovery.

F_{MAX}/V_{MIN} and noise margin of the full SRAM array over its lifetime are governed by the combination of WID parameter variations at BOL and statistical aging impacts from BOL to EOL across all bits in the array. Read F_{MAX} for target 1Mb array improves by 60 (40) MHz for idle (active) mode stress with corresponding 25 (21) % noise margin degradation (Fig. 4). Thus, aging-aware dynamic keeper upsizing from BOL to EOL can achieve higher read F_{MAX} while maintaining adequate noise margin over the operational lifetime. Write-1 (write-0) V_{MIN} at EOL improves by 20/10 (30/20) mV for idle/active mode stress. Retention-1 V_{MIN} degrades by 20mV at EOL due to weakening of the holding PMOS/NMOS in the cell and the write-assist PMOS (Fig. 6). Retention-0 V_{MIN} first lowers by 30mV, but ends up unchanged at EOL due to weakened write-assist PMOS. Thus, aging aware adaptive voltage scaling can reduce total energy consumption in active/idle modes over the lifetime.

Acknowledgements

This research was, in part, funded by the U.S. government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the U.S. government.

References

[1] T.-H. Kim, et al., *JSSC,* Apr 2008*,* pp. 874-880
[2] M. Igarashi, et al., ESSCIRC 2015, pp. 112-115
[3] M. Cho, et al., CICC 2015, pp. 1-4
[4] M. Yuffe, et al., ISSCC 2011, pp. 264-266

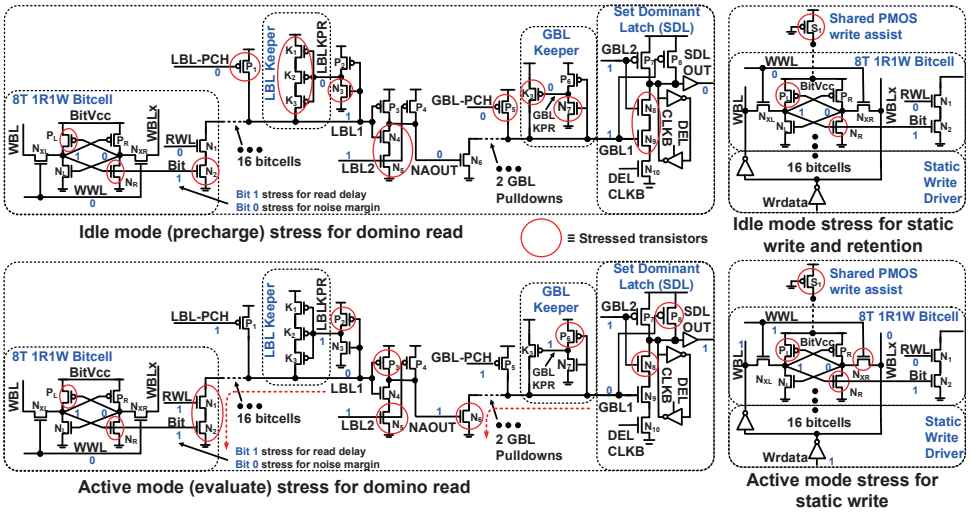
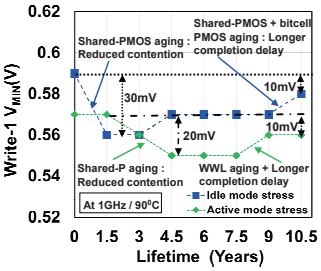


Fig. 1 Idle/active mode stress scenarios for domino-read, static-write and retention in 1R1W 8T SRAM array



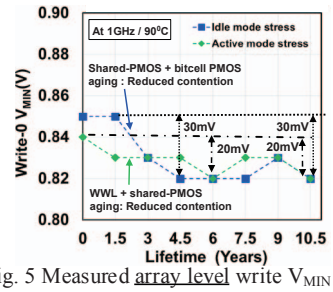
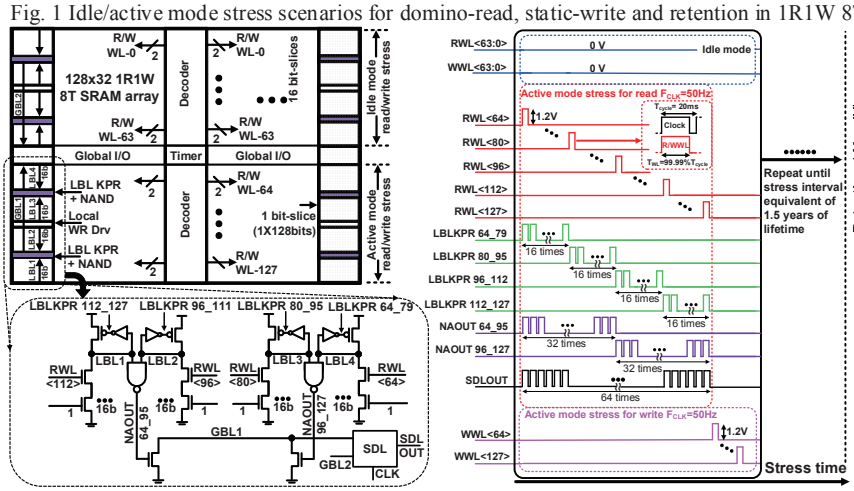


Fig. 5 Measured array level write V_{MIN} for 1Mb target over operational lifetime



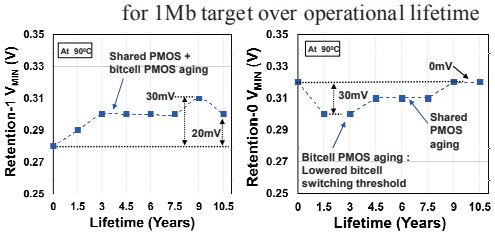
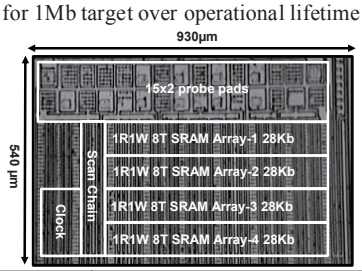
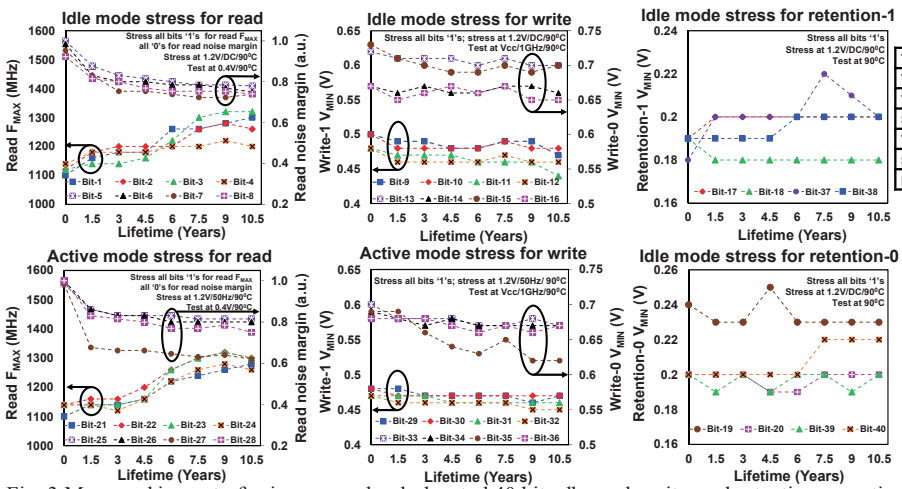


Fig. 6 Measured array level retention V_{MIN} for 1Mb target over operational lifetime



Technology	22nm High-K/Metal-Gate tri-gate CMOS
Total bit count	112 Kb
Transistors	~1.5 million
Stress Condition	1.2V/90°C
Total Stress Duration	Equivalent to 10.5 years of lifetime
Stress Interval	Equivalent to 1.5 years of lifetime
Nominal Condition	Vcc= up to 0.8V, Temp= 90°C, Max Freq.= 1.6 GHz



Parameter	Aging Impact on 8T SRAM Array@ EOL	
	Idle mode stress	Active mode stress
Read F_{MAX}	60MHz ↑	40MHz ↑
Read noise margin	25% ↓	21% ↓
Write-1 V_{MIN}	20mV ↓	10mV ↓
Write-0 V_{MIN}	30mV ↓	20mV ↓
Retention-1 V_{MIN}	20mV ↑	---
Retention-0 V_{MIN}	0mV	---

Fig. 7: 22nm high-k/metal gate tri-gate CMOS test-chip micrograph and measurement summary table showing idle and active mode aging impact at EOL on 1R1W 8T SRAM array

Fig. 3 Measured impact of aging on randomly located 40 bitcells read, write, and retention operations: bits 1-20 are located in the lower half array (idle mode), bits 21-40 are located in the upper half (active mode) array

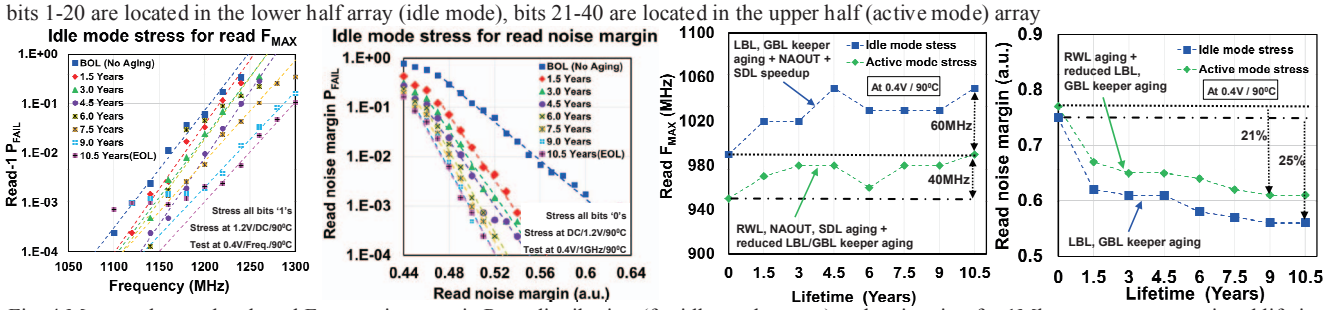


Fig. 4 Measured array level read F_{MAX}, noise margin P_{FAIL} distribution (for idle mode stress) and estimation for 1Mb target over operational lifetime