Characterization of Dynamic SRAM Stability in 45 nm CMOS

Seng Oon Toh, *Student Member, IEEE*, Zheng Guo, *Member, IEEE*, Tsu-Jae King Liu, *Fellow, IEEE*, and Borivoje Nikolić, *Senior Member, IEEE*

***Abstract*—Optimization of SRAM yield using dynamic stability metrics has been evaluated in the past to ensure continued scaling of bitcell size and supply voltage in future technology nodes. Various dynamic stability metrics have been proposed but they have not been used in practical failure analysis and compared with conventional static margins. This work compares static and dynamic metrics to identify expected correlations. A dynamic stability characterization architecture using pulsed word-lines is implemented in 45 nm CMOS to identify sources of variability, and their impact on SRAM stability. Static read margins were observed to overestimate failures by 10–100 X while static write margins failed to predict outliers in critical writeability. Critical writeability was demonstrated to exhibit an enhanced sensitivity to process variations, random telegraph noise (RTN), and negative bias temperature instability (NBTI), compared to static write margins.**

***Index Terms*—Dynamic stability, NBTI, pulsed word-line, RTN, SRAM, variability.**

I. Introduction

SRAM scaling has been identified as one of the bottlenecks for supply voltage (V_{DD}) reduction in current and future technology nodes. Minimum SRAM operating voltage (V_{MIN}) is a function of the magnitude of process-induced variability as well as array size. Aggressive SRAM bitcell scaling, as well as continued increase in SRAM array sizes, has resulted in stagnation in SRAM V_{DD} scaling. This trend is observed in reported values of SRAM array V_{DD} and is recognized in the latest edition of the International Technology Roadmap for Semiconductors (ITRS) (Fig. 1) [1]. V_{MIN} is traditionally estimated using static margins such as static noise margin (SNM) and N-curves [2], [3]. These metrics are known to be optimistic in writeability and pessimistic in read stability from comparisons between static and actual dynamic access [26].

Manuscript received March 27, 2011; revised July 11, 2011; accepted July 11, 2011. Date of publication September 19, 2011; date of current version October 26, 2011. This paper was approved by Associate Editor Peter Gillingham. This work was supported by the Center for Circuit & System Solutions (C2S2) Focus Center, one of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program.

S. O. Toh is with Advanced Micro Devices, Sunnyvale, CA 94085 USA and also with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94704 USA (e-mail: sengoon@eecs.berkeley.edu).

Z. Guo is with Intel Corporation, Hillsboro, OR 97124 USA.

T.-J. K. Liu and B. Nikolić are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94704 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2011.2164300

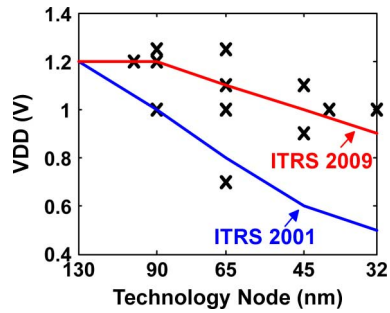


Fig. 1. SRAM array V_{DD} reported in ISSCC and VLSI (2004–2010) and ITRS predictions.

Dynamic stability metrics, derived from the SRAM under dynamic access, have been proposed to provide a better estimate of SRAM V_{MIN} [4]–[6]. While these metrics have been studied extensively through simulations, results based on large-scale silicon characterization of both read and write stability have not yet been reported. Similarly, a quantitative relationship between the static and dynamic read and write margins has not been studied. The sensitivity of dynamic stability to non-idealities such as random telegraph noise (RTN) and aging is still largely an open problem.

In this work, we propose a characterization architecture for measuring dynamic SRAM stability through pulsed word-lines calibrated up to 10 ps accuracy [7]. Measuring word-line pulse-widths calibrates out any timing uncertainty introduced by SRAM peripheral circuits, thus allowing characterization of the fundamental variability of the SRAM bitcells. This characterization methodology is validated in a commercial low-power 45 nm CMOS process. The test chip also provides a means of correlation with static read and write metrics via direct bit-line measurements [8]. This method is used to identify new sources of variability in dynamic stability by observing deviations from expected correlations between dynamic stability and static margins.

We first review conventional static and dynamic 6 transistor SRAM metrics as well as their expected correlations in Section II. Monte Carlo simulations, introducing Gaussian distributions of V_{th} to the 6 SRAM transistors, are presented in this section to illustrate expected correlations between the metrics. Section III presents the proposed dynamic stability characterization architecture while Section IV describes an implementation in a 45 nm CMOS test chip. Section V summarizes measurement results and their implications. Finally, conclusions are given in Section VI. All voltage margins in the text are normalized to the supply voltage. Studied margins

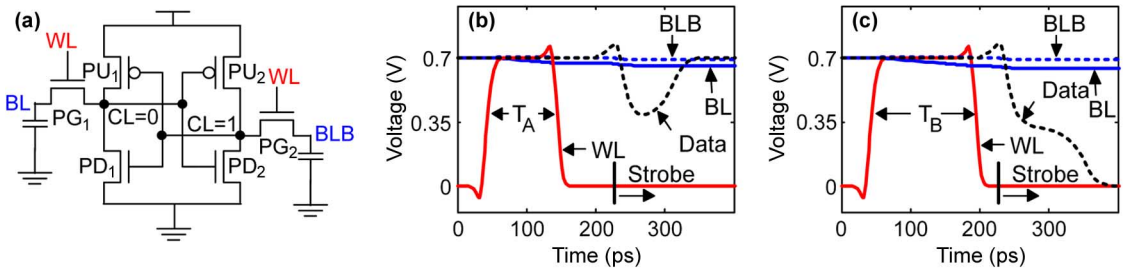


Fig. 2. (a) Schematic of a 6-T SRAM cell storing a "0" on the left internal node. (b) Simulated waveforms corresponding to failed read access with pulse-width, T_A. Output of the sense-amplifier (*Data*) resolves to the incorrect value. (c) Simulated waveforms corresponding to successful read access with a longer pulse-width, T_B. Output of the sense-amplifier (*Data*) resolves to the correct value.

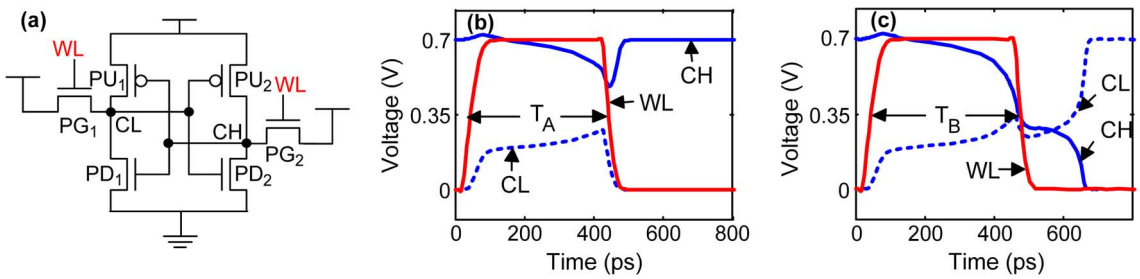


Fig. 3. (a) Schematic of a 6-T SRAM cell under read stress. (b) Simulated waveforms corresponding to read stable access with pulse-width, T_A. The state of the bitcell is retained after read operation. (c) Simulated waveforms corresponding to read upset with a longer pulse-width, T_B. The state of the bitcell is accidently flipped by the read operation.

typically exhibit proportionality to the supply voltage, and normalizing them allows for comparison with prior studies (e.g. [8], [9]).

II. Static and Dynamic SRAM Metrics

A. Read Access

1) Static Read Current ($I_{\rm read}$): $I_{\rm read}$ corresponds to the current that is being sourced from the bit-line into the SRAM node storing a "0". Under SRAM read operation, this current is responsible for discharging the pre-charged bit-line capacitances ($C_{\rm BL}$) enough to overcome the offset voltage ($V_{\rm offset}$) of the sense-amplifier to result in a correct value being latched. It is expected to correlate with actual read access time ($T_{\rm access}$):

$$T_{\rm access} \propto \frac{C_{\rm BL} \times V_{\rm offset}}{I_{\rm read}}. \quad (1)$$

Actual read access time might deviate from this linear relationship due to leakage currents from inactive bitcells sharing the bit-line as well as the fact that $C_{\rm BL}$ is a distributed RC network spanning the entire column of the SRAM array. Degradation in $I_{\rm read}$ due to RTN also contributes to this discrepancy, as will be shown in Section V.

2) Read Access Time ($T_{\rm access}$): Fig. 2 illustrates an SRAM bitcell undergoing read access with pulse-widths T_A and T_B. Pulse-width T_A is too short to sufficiently discharge the bit-line capacitance to overcome offset in the sense-amplifier. There exists a critical pulse-width, $T_{\rm access}(T_A < T_{\rm access} < T_B)$, where the sense-amplifier is on the threshold of a successful read access that is defined as the read access time. This is similar to the dynamic access failure criteria defined in [5]. This definition of read access time isolates out variability in the read access operation due to variability of the SRAM bitcell and ignores other delays such as word-line driver delay and sense-amplifier delay.

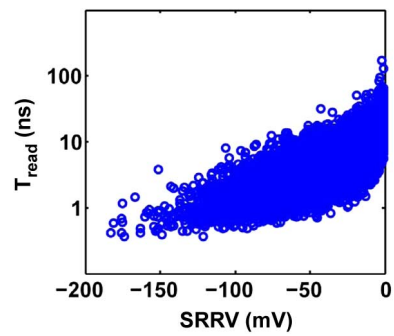


Fig. 4. Simulated scatter plot showing the correlation between critical read stability ($T_{\rm read}$) and negative static read margin (SRRV).

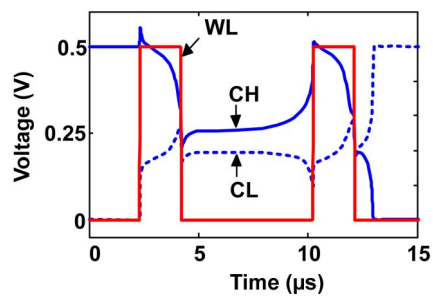


Fig. 5. Simulated waveforms corresponding to an SRAM bitcell under read-after-read access.

B. Read Stability

1) Static Read Stability Margins: Conventional stability metrics, such as SNM and N-curves [2], [3], require sweeping internal nodes in order to obtain the voltage transfer curves, which is not practical for evaluating large arrays. We choose to characterize the supply read retention voltage (SRRV), which does not require access to the internal nodes. A direct correlation between

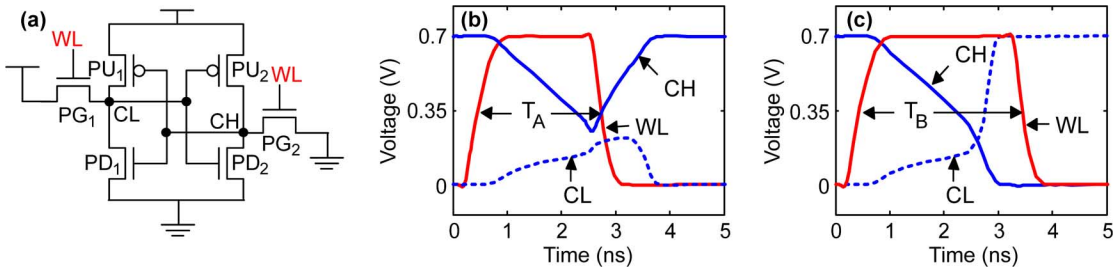


Fig. 6. (a) Schematic of a 6-T SRAM cell under write access. (b) Simulated waveforms corresponding to failed write access with pulse-width, T_A. The bitcell retains original state. (c) Simulated waveforms corresponding to successful write access with a longer pulse-width, T_B. A new value is written into the bitcell.

this and other stability metrics has already been established in [8].

2) Critical Read Stability ($T_{\rm read}$): Fig. 3 illustrates an SRAM bitcell undergoing read stress with pulse-widths T_A and T_B. Pulse-width T_A is short enough that the internal nodes (CH and CL) return back to their original levels after the word-line pulse. The longer pulse-width T_B subjects the bitcell to too much read stress, causing the cell to flip to an opposite state after the word-line pulse. There exists a critical pulse-width, $T_{\rm read}$ ($T_A < T_{\rm read} < T_B$), where the bitcell is on the threshold of a read upset, that is defined as the critical read stability. This is similar to the dynamic read failure criteria defined in [5]. This metric does not require access to the internal nodes of the SRAM cell. The challenge is to reliably evaluate the contents of the bitcell after the test, without accidently disrupting the stored state.

A bitcell with positive static read margin will have infinite $T_{\rm read}$ while a bitcell with zero or negative static read margin will have a finite value of $T_{\rm read}$. With the SRRV margin, it is possible to characterize a negative static read margin for a particular bitcell by measuring how much additional bitcell $\rm V_{DD}$ ($V_{\rm CELL}$), above the nominal voltage, is required to maintain the stored state of the SRAM cell. Increasing the $V_{\rm CELL}$ of a statically unstable bitcell by the absolute value of its negative SRRV, results in infinite $T_{\rm read}$. Fig. 4 plots the positive correlation observed between SRRV and $T_{\rm read}$ extracted from Monte Carlo simulations. Although $T_{\rm read}$ is observed to be exponentially dependent on static read margin, it is impossible to accurately estimate exact values of critical read stability from a voltage screen test at elevated $V_{\rm CELL}$ due to the large dispersion (up to 10x) observed in $T_{\rm read}$ at a particular SRRV.

SRAM access with read-after-read operation presents the worst-case condition for critical read stability [5], [6]. Fig. 5 illustrates the waveforms corresponding to an SRAM bitcell with read-after-read access. The SRAM bitcell is stable after the first word-line pulse but is subsequently corrupted by the second pulse. It is therefore important to characterize $T_{\rm read}$ as a function of the number of read-after-read pulses as well as the access frequency.

C. Writeability

1) Static Writeability Margins: Margins such as write noise margin (WNM) and write N-curve require sweeping internal nodes in order to obtain the voltage transfer curves [9], [10]. We choose to characterize bit-line write trip voltage (BWTV) that can be measured by sweeping the bit-line voltages of the SRAM bitcell. Correlations established with this margin can be extended to other static margins based on previously established relationships [8].

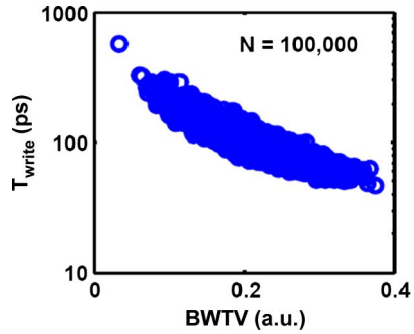


Fig. 7. Simulated scatter plot comparing critical writeability ($T_{\rm write}$) and static write margin (BWTV) obtained from Monte Carlo simulations.

2) Critical Writeability ($T_{\rm write}$): Fig. 6 illustrates write operation to a SRAM bitcell with pulse-widths T_A and T_B. Pulse-width T_A is too short to overwrite the contents of the SRAM cell while pulse-width T_B is sufficient to complete the write operation. There exists a critical pulse-width, $T_{\rm write}$ ($T_A < T_{\rm write} < T_B$), where the bitcell is on the threshold of a successful write access that is defined as the critical writeability. This is similar to the dynamic write failure criteria defined in [5]. This metric does not require access to the internal nodes of the SRAM cell. The challenge, however, is to reliably evaluate the contents of the bitcell after the test, without accidently disrupting the stored state.

Fig. 7 plots the expected correlation between $T_{\rm write}$ and static write margin, based on Monte Carlo simulations. Bitcells with poor static write margin (smaller values) are expected to be correlated with poor $T_{\rm write}$ (larger values). The dispersion between $T_{\rm write}$ and BWTV is small, especially at lower static margins, implying the possibility of using voltage screening either by reducing $V_{\rm CELL}$ or word-line bias to identify cells with poor $T_{\rm write}$. Table I tabulates the sensitivities between the respective write margins to $V_{\rm th}$ variability in the 6 transistors of an SRAM bitcell under write operation as illustrated in Fig. 6(a). The sensitivities in Table I reflect the negative correlation between BWTV and $T_{\rm write}$. Both margins have similar magnitude of sensitivities except for the pull-up transistors, as $T_{\rm write}$ is correlated with variability in transistor $PU1$ while BWTV is independent, and $PU1$ is positively correlated with poor $T_{\rm write}$

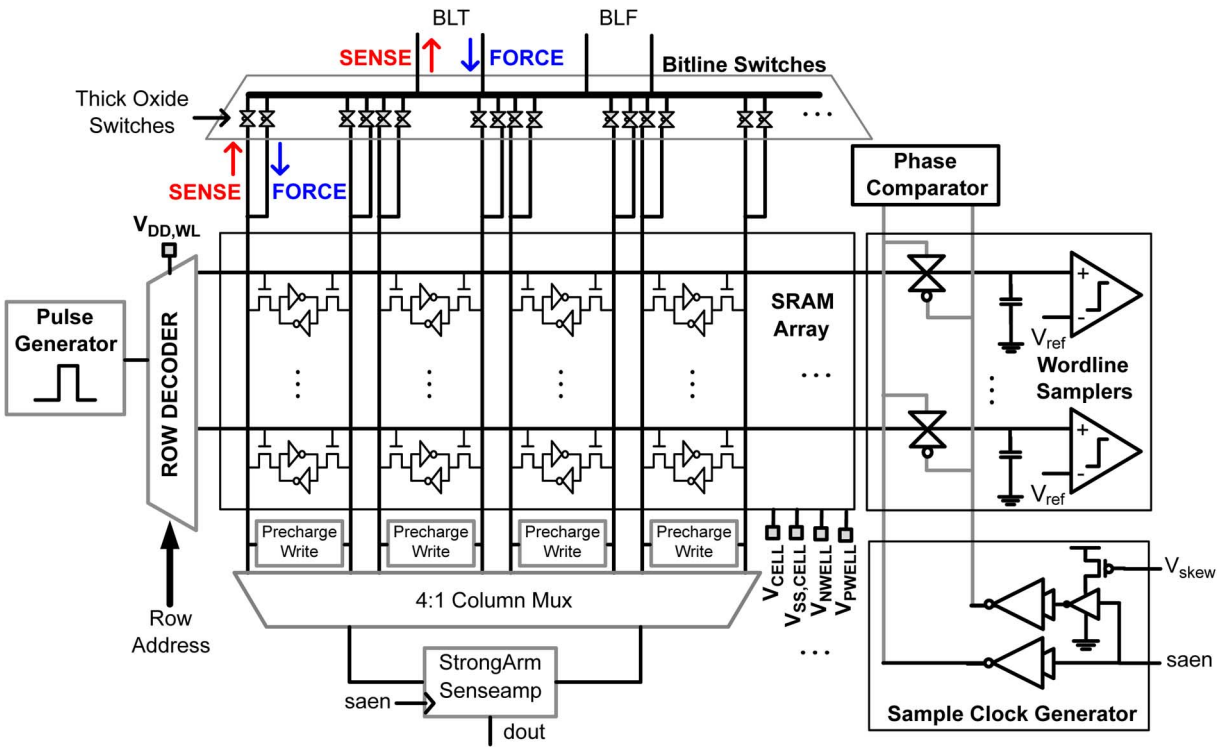


Fig. 8. SRAM array organization for static and dynamic stability characterization.

TABLE I
SENSITIVITY ANALYSIS OF WRITEABILITY TO THE RESPECTIVE TRANSISTOR V_{th} VARIATION

	PD1	PG1	PU1	PD2	PG2	PU2
BWTV (V/V)	0.3	-0.3	0	0	-0.8	0.6
T_{write} (ns/V)	-0.08	1.0	1.3	0.02	3.9	-0.7

while $PU2$ is negatively correlated with poor $T_{\rm write}$. This suggests that $T_{\rm write}$ is more susceptible to cell asymmetry than static write margin. Read-before-write or read-after-write does not need to be considered because the read operation only helps to upset the cell and complete the write operation [5]. $T_{\rm write}$ under write-after-write access, however, needs to be characterized to evaluate the impact of RTN on $T_{\rm write}$.

III. Dynamic Stability Characterization Architecture

Fig. 8 presents the SRAM array configuration for the characterization of dynamic metrics. It also shows the necessary infrastructure for collecting static metrics for the purpose of establishing correlations with dynamic metrics. The SRAM bitcells under test are organized into a conventional SRAM array. Various array bias voltages ($V_{\rm DD,WL}$, $V_{\rm CELL}$, $V_{\rm SS,CELL}$, $V_{\rm NWELL}$, and $V_{\rm PWELL}$) are connected to pads to characterize the SRAM under different read/write assist modes. A programmable pulse is generated on-chip and delivered to a single word-line at a time using existing row decoders. This architecture makes extensive use of simple circuits and calibration to ensure ease of implementation while providing measurements with high fidelity even in highly-scaled process technologies. A programmable pulse is generated by simply mixing together two clocks, Φ_0 and Φ_1, that have a slight offset in clock period (ΔT) (Fig. 9). This generates a pulse train with a difference in pulse-width of ΔT between successive pulses. A counter is then used to pass the desired pulse based on a programmed codeword. This pass signal can also be programmed to be held for multiple clock cycles to generate multiple pulses, simulating read-after-read access. The *sync* signal used to reset the counter is generated digitally on-chip based on statistics of the beat frequency between Φ_0 and Φ_1, averaged over 128 samples to minimize the impact of clock jitter.

To avoid process-induced uncertainties, the exact pulse width is measured by word-line samplers located on every word-line (Fig. 8). This contrasts to prior work in which a small subset of the word-lines is sampled [11], [12]. The sampler consists of small transmission gates sampling the word-line pulse on a parasitic capacitance. Charge injection by the sampling clock, non-linearity of the transmission gates, and offset voltages of the comparators are calibrated out by tuning the reference voltage of the comparators. The differential clock driving the transmission gates is calibrated using a phase comparator to minimize aperture uncertainty in sampling the rising and falling edges of the word-line pulse (Fig. 10). An ideal differential clock should have no common mode component. This phase comparator takes advantage of this fact and detects the common mode component by summing these two signals using capacitors. The calibration scheme then proceeds to skew the edges of the clock until the glitch on the sum node is minimized. A Monte Carlo simulation of this scheme reveals that it reduces the phase offset of respective edges to less than 3 ps. The word-line

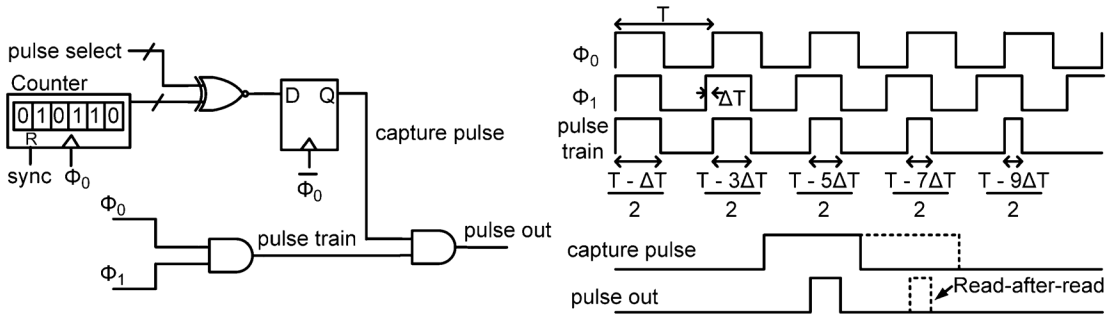


Fig. 9. Frequency mixing programmable pulse generator with corresponding waveforms.

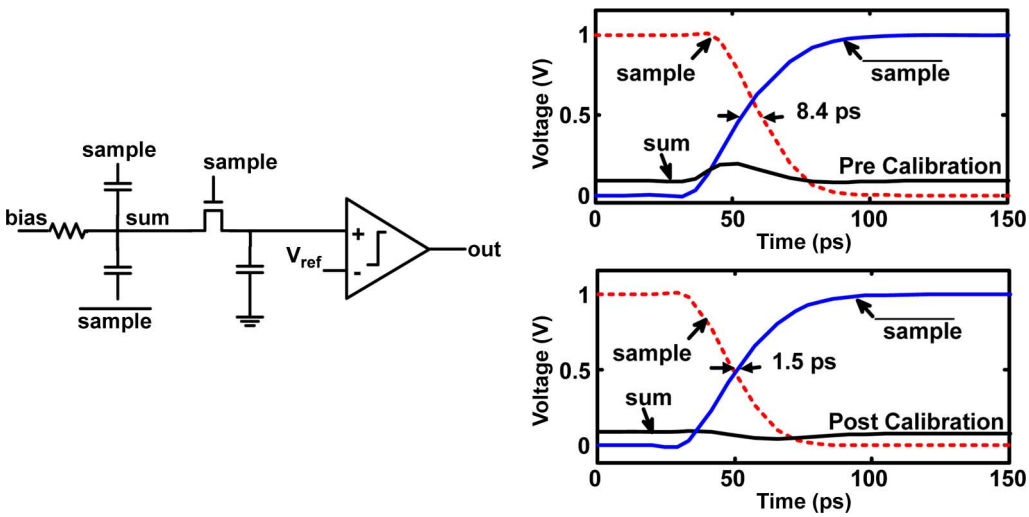


Fig. 10. Capacitive summing phase comparator and simulated waveforms before and after calibration.

pulse-width is finally measured by skewing the externally generated *saen* signal with respect to Φ_0 with 1 ps resolution. This word-line sampling scheme produces finer resolution compared to delay-line samplers [11].

Non-destructive read-back of the SRAM bitcells is accomplished using multiple minimum-width read pulses. This allows the bitcell to gradually discharge the bitline capacitance without excessive read stress. Alternatively, V_{CELL} is raised to the nominal voltage prior to read-back, especially when characterizing bitcells at low voltages. A built-in-self-test (BIST) circuit is used to characterize the dynamic stability of each bitcell automatically. The static margins of the SRAM bitcells are measured through the bit-lines using source meters with four-terminal Kelvin sensing to calibrate out the series resistance of the bit-line switches [8]. I–V characteristics and RTN in each individual transistor of a 6T SRAM bitcell were characterized using the direct bit transistor access (DBTA) method [24].

IV. 45 nm CMOS Test Chip

A 1.55 mm $\times$ 1.55 mm test chip [7], [13], [14] is implemented (Fig. 11) in a low-power strained-Si 45 nm CMOS process [15] with poly- Si/SiO_xN_y gate stack and seven metal layers. Experimental, high density 0.252 μm^2 6T SRAM bitcells that are smaller than ITRS requirements for the 45 nm technology node are characterized to observe a larger impact of process-induced variability on SRAM performance and also to predict variability in future scaled transistors. The test chip consists of two 64 $\times$ 256 arrays and two 128 $\times$ 256 arrays with full static and dynamic stability characterization coverage. The narrower array (64 columns) has reduced word-line parasitics and is used to characterize dynamic stability at high speeds with strict requirements of rise- and fall-transition times. The word-line samplers contribute to a 16% array area overhead. The level-shifters and bit-line switches incur a larger area penalty and are required solely for static margin characterization.

V. Measurement Results

Fig. 12 illustrates fail bit count measured from the test chip, indicating 10–100X discrepancy between quasi-static (>1 s with bit-lines driven) and dynamic access. Static access fail bit counts are optimistic for writeability and pessimistic for read stability, compared to those for dynamic access. More than 10 write failures were observed at nominal V_{DD} when the bitcells were accessed with 1 ns pulses even though no write failures occurred when the bitcells were accessed quasi-statically. No read upset failures occurred when the bitcells were accessed with 20 ns pulses even though tens of failed bits were observed when the same bitcells were accessed quasi-statically.

A. Pulse Generator

Multiple complete waveforms of word-line pulses were subsampled and plotted in real time in Fig. 13(a). Good rise and fall transition times of 75 ps and 30 ps were observed. Note that

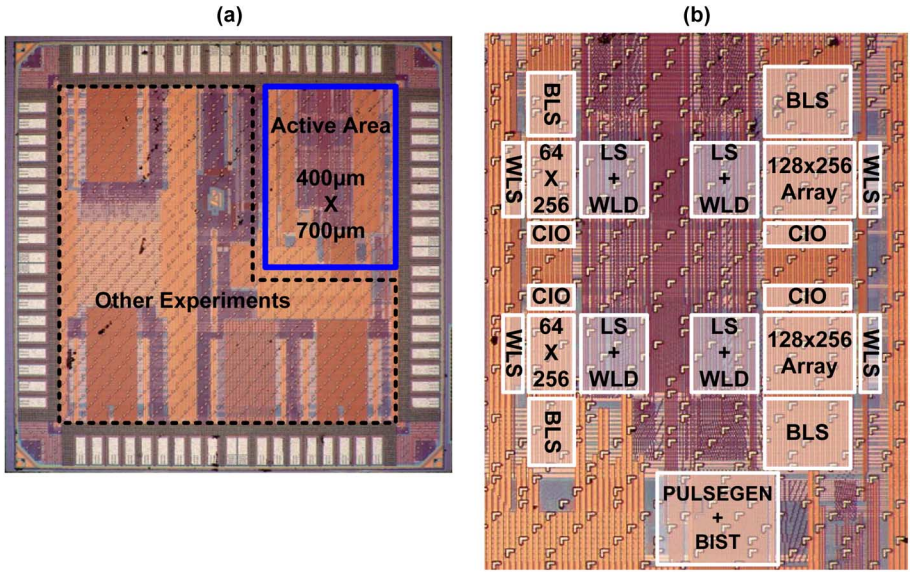


Fig. 11. (a) Die photo of the 45 nm CMOS test chip. (b) Die photo of active area. BLS: bitline switches; WLS: word-line samplers; LS+WLD: level shifters and word-line drivers; CIO: column I/O circuitry.

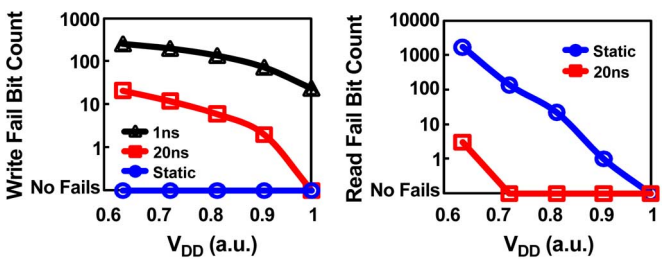


Fig. 12. SRAM writeability and read stability fail bit count measured from a 45 nm CMOS SRAM.

the rise and fall transitions account for a significant portion of narrow pulses (less than 100 ps) and effectively limit the correlation between static and dynamic margins. The pulse-width, corresponding to the delay between the 50% voltage level of the rise and fall transitions, was measured across different codewords. The transfer function and the measured linearity error are plotted in Fig. 13(b). Up to 100 ps of non-linearity was observed in the transfer function. This error is believed to be caused by voltage droop in the power supply grid as the pulse is being distributed across the chip. These non-idealities demonstrate the importance of calibrating word-line pulse-widths at every word-line in order to calibrate out this source of uncertainty from actual variability in the bitcells. All dynamic SRAM measurements presented are based on word-lines calibrated to 10 ps resolution using low-jitter signal generators and averaging.

B. Read Access Time

Fig. 14(a) plots the statistical distribution of $T_{\rm access}$ measured from 1024 bitcells at 0.8X nominal $V_{\rm DD}$. The distribution is observed to be multi-modal, a superposition of multiple Gaussian distributions. The multi-modal nature of this distribution is due to the strong dependence of read access time on sense-amplifier offset voltage, (1). Measurements of $T_{\rm read}$, normalized with separately characterized sense-amplifier offset voltages and estimated bit-line capacitance, was observed to correlate ($R^2 = 0.69$) with static read current (Fig. 14(b)). The remaining dispersion in the data is due to the inherent difference between $I_{\rm read}$ statically measured out of the bitcell at a fixed bit-line voltage and the transient bitcell current as the bit-line is being discharged.

C. Critical Writeability

Fig. 15(a) plots measurements of critical writeability versus the static write margin for writing the same data value to the same bitcell. Each data-point of $T_{\rm write}$ corresponds to an average of 128 measurements. Expected correlation between poor BWTV and $T_{\rm write}$ is observed in Fig. 15(a), however, the uncorrelated outliers exceed the correlated data-points by more than ten times. These outliers are observed to appear exclusively in bitcells that have large static write margin on the opposite side of the cell (Fig. 15(b)). Further analysis of individual transistor characteristics using DBTA revealed that a large number of bitcells sampled had large drain series resistance in one of the PMOS transistors. These marginal transistors were found to be on the side opposite to the half-cell being written to ($PU1$ in Fig. 6(a)), causing a significant degradation in the speed of the bitcell for pulling the storage node up to $\rm V_{DD}$. The remaining bitcells showed good correlation between $T_{\rm write}$ and BWTV metrics, after the marginal cells were screened out (Fig. 15(a)). These marginal transistors did not degrade static write margin due to the negligible sensitivity of the margin to variability in $PU1$ (Table I).

Voltage screen tests such as described in [16] are commonly used to screen out defects and early failures in SRAM arrays. Such tests are usually carried out in-line at wafer sort using testers running at lower frequencies than actual operating frequencies. The lack of correlation between the outliers in critical writeability and static write margin invalidates results obtained

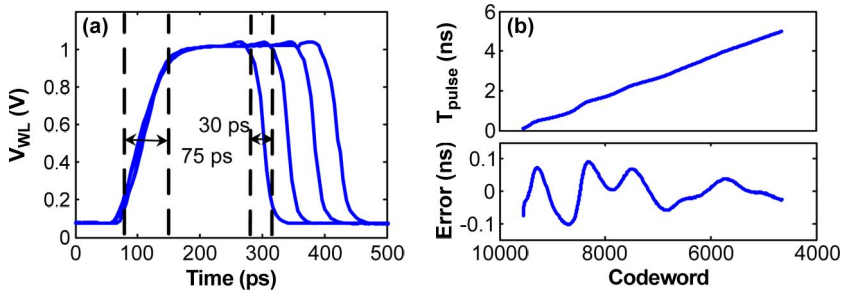


Fig. 13. Plots of (a) multiple sub-sampled word-line waveforms and (b) codeword to pulse width transfer function and measured error.

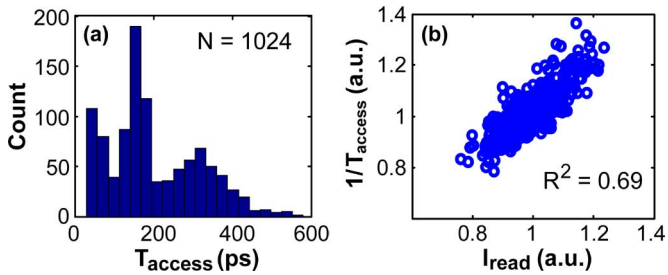


Fig. 14. (a) Histogram of measured read access time. (b) Scatter plot showing correlation between read access time and static read current after normalization with sense-amplifier offset voltage and bit-line capacitance.

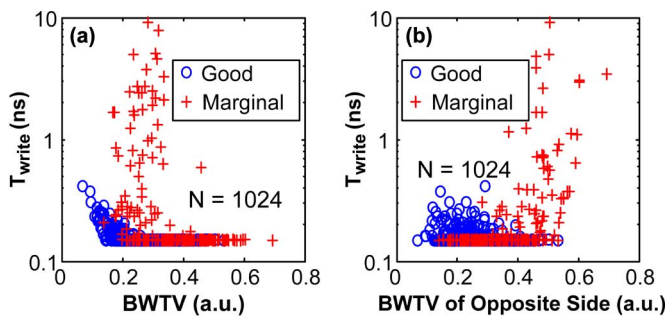


Fig. 15. Critical writeability versus static write margin of (a) same side and (b) opposite side of SRAM cell measured at $V_{DD,low}$.

from such tests because the bitcells screened by these tests are not the bitcells that fail first at normal operating frequencies.

D. Critical Read Stability

Fig. 16 plots measurements of critical read stability against the negative static read margin. These measurements were obtained by lowering $V_{\rm CELL}$ by 300 mV relative to word-line and bit-line pre-charge voltage levels, to increase the probability of observing cells that are unstable under static access. The expected correlation between $T_{\rm read}$ and negative SRRV (ref. Fig. 4) was observed in measurements. Bitcells with marginally negative static read margin (approximately 0.1 a.u.) were observed to have a large dispersion in $T_{\rm read}$ ranging from 1 ns to 1 μs. This dispersion reduces as the bitcell SRRV becomes more negative. The minimum $T_{\rm read}$ observed was 630 ps, indicating that this SRAM bitcell can be accessed with pulse-widths shorter than 630 ps without read upsets even with 300 mV of $V_{\rm CELL}$ droop. Outliers with extremely poor SRRV that are not

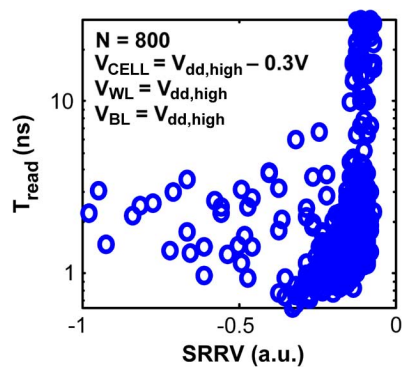


Fig. 16. Critical read stability versus static read margin.

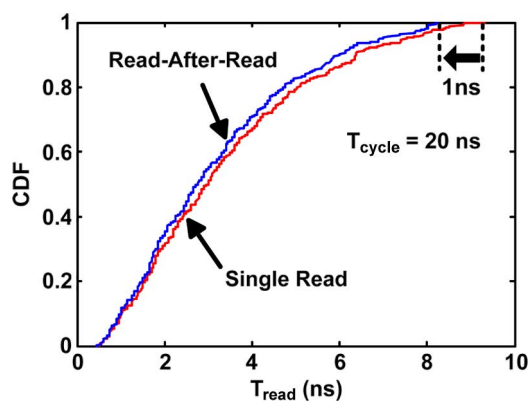


Fig. 17. Statistical distributions of critical read stability under single read and read-after-read access with 20 ns clock period.

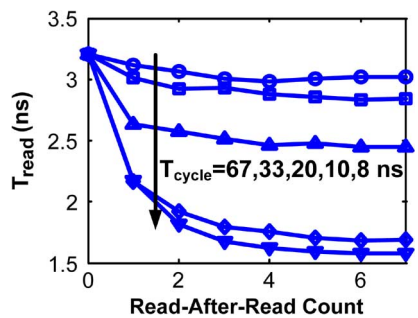


Fig. 18. Critical read stability of a selected bitcell as a function of the number of read-after-read cycles. The different curves correspond to the period of the read-after-read cycles.

correlated with smaller values of $T_{\rm read}$ were observed. Such outliers were not observed in Monte Carlo simulations of a large 100,000 sample set (Fig. 4).

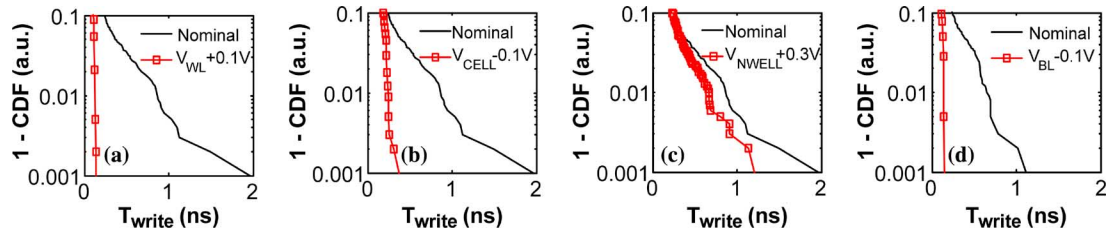


Fig. 19. Survival function of $T_{\rm write}$ under different bias conditions: (a) Word-line boosting; (b) $V_{\rm CELL}$ under-drive; (c) PMOS reverse body-bias; (d) Negative bit-line.

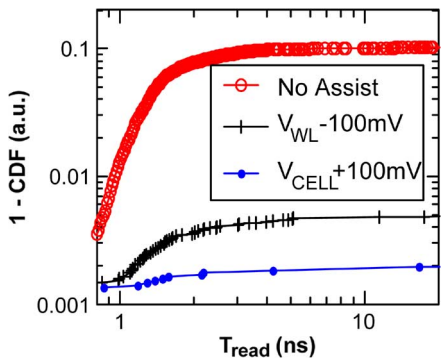


Fig. 20. Survival function of $T_{\rm read}$ without assist techniques, with -100 mV word-line bias offset and with $+100$ mV $V_{\rm CELL}$ offset.

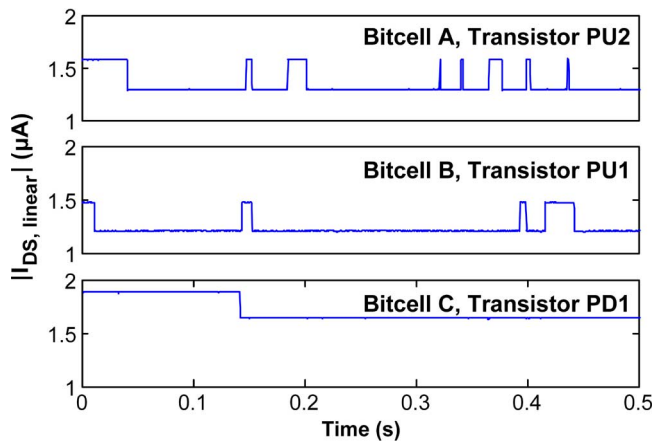


Fig. 21. Drain current with RTN measured from three transistors in three different bitcell instances.

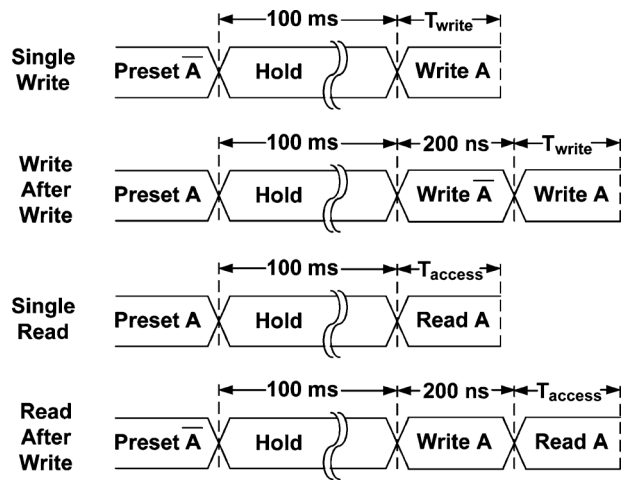


Fig. 22. Different SRAM access patterns for evaluating the impact of RTN on dynamic stability.

Fig. 17 plots statistical distributions of critical read stability under single read and read-after-read access with 20 ns period $(T_{\rm cycle})$. As expected, $T_{\rm read}$ degrades under read-after-read conditions [5]. Bitcells with small values of $T_{\rm read}$ (less than 2 ns) were observed to shift only by a small amount, while bitcells with larger $T_{\rm read}$ were observed to degrade by up to 1 ns. Susceptibility of a bitcell to read-after-read upset depends on the proximity of the internal node voltages to the rails when the next read pulse arrives. Bitcells with smaller values of $T_{\rm read}$ are less susceptible to read-after-read upsets, compared to bitcells with larger $T_{\rm read}$ accessed with the same $T_{\rm cycle}$, because these bitcells have longer recovery periods to settle at the rail voltages. Fig. 18 plots $T_{\rm read}$ of a single bitcell as a function of the number of read-after-read pulses across decreasing $T_{\rm cycle}$. The degradation in $T_{\rm read}$, due to read-after-read, increases as $T_{\rm cycle}$ is decreased. This degradation saturates eventually after 6 cycles in direct agreement with [5]. Evidence of slight $T_{\rm read}$ degradation even with a relatively slow $T_{\rm cycle}$ of 67 ns suggests that the recovery period of this bitcell is more than 67 ns, which is greater than 20 times the single-read $T_{\rm read}$ of this bitcell (3.2 ns).

E. Impact of Assist Techniques

Fig. 19 compares the impact of different assist techniques on $T_{\rm write}$. Word-line voltage $(V_{\rm WL})$ boosting and $V_{\rm CELL}$ under-drive resulted in significant speed-up of $T_{\rm write}$ [28]. $V_{\rm WL}$ boost was slightly more effective than $V_{\rm CELL}$ under-drive because it increases the strength of the pass-gate transistors which have the strongest impact on $T_{\rm write}$. Fig. 19(c) plots the statistical distributions of $T_{\rm write}$ under 300 mV of PMOS reverse body-bias (RBB) [29]. Not much improvement in $T_{\rm write}$ was observed even with 300 mV of RBB due to the small body-effect coefficient for this 45 nm CMOS process. RBB might even have a detrimental effect on $T_{\rm write}$, due to the opposite sensitivities of $T_{\rm write}$ to variability in $PU1$ and $PU2$. Fig. 19(d) investigates write assist using negative voltage levels on the bit-lines [30]. A 100 mV negative bit-line bias results in a significant improvement in $T_{\rm write}$.

Fig. 20 demonstrates the effectiveness of $V_{\rm CELL}$ boosting and $V_{\rm WL}$ under-drive for read assist [17]. $V_{\rm CELL}$ boosting was found to provide a larger improvement in critical read stability compared to $V_{\rm WL}$ under-drive. SRAM design using assist techniques involves a delicate balance of bias voltages in order to balance out the improvement in one margin with the degradation

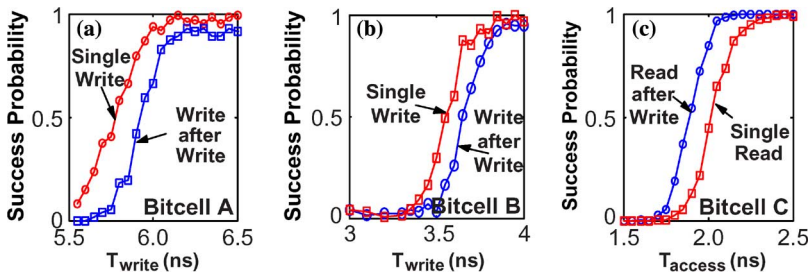


Fig. 23. Statistical distributions corresponding to: (a) writing $CH = 0$ into bitcell A, (b) writing $CH = 0$ into bitcell B, (c) reading $CL = 0$ from bitcell C.

in the other. The strong sensitivity of read and write stability to $V_{\rm WL}$ and $V_{\rm CELL}$ biasing suggests the possibility of using these two voltage tuning knobs to increase the overall reliability of the SRAM array. Results in this work however demonstrate that this technique needs to be used with caution as slight offsets in $V_{\rm WL}$ will affect $T_{\rm write}$ and $T_{\rm read}$ exponentially. Because of this, any uncertainty or noise in setting $V_{\rm WL}$ can result in large write or read stability failures.

F. Impact of Temporal Variations

1) Random Telegraph Noise (RTN): RTN refers to a noise phenomenon that is caused by charge trapping and de-trapping within the gate oxide of the transistor [21]. Aggressive scaling of SRAM transistor active area has resulted in an increasing contribution of RTN to transistor variability compared to random dopant fluctuations [18]. While RTN is observed in SRAM operation as low frequency fluctuation in static read and write margins, the impact of RTN on SRAM operating at high frequency has not yet been evaluated [19], [20].

Fig. 21 plots drain currents of transistors measured from three different bitcells using the DBTA method. The pass-gate transistors were biased into strong inversion to reduce the contribution of RTN in the pass-gates to the measured drain current, as the pass-gate transistors are in series with these transistors. These bitcells were selected because the RTN amplitude fluctuation from the selected transistors was much larger than the other transistors in the same bitcell. This allowed direct correlation between characteristics observed in dynamic stability metrics and RTN in a particular transistor. Dynamic stability of the bitcells was characterized with different dynamic access patterns (Fig. 22), designed to emphasize the impact of RTN on dynamic stability.

Fig. 23 plots statistical distributions extracted from the respective access patterns on the corresponding bitcells, averaged over 128 tries at each pulse-width. Low frequency RTN in the transistors resulted in shifts in bitcell dynamic stability of up to 11%, that is dependent on single or multiple access [27]. Write-after-write access degraded $T_{\rm write}$ corresponding to writing $CH = 0$ into both bitcell A and bitcell B even though large RTN was observed in different transistors ($PU2$ and $PU1$). This shift effect can be explained by considering the large-signal dependence of RTN trap occupancy [25]. The 100 ms $CH = 0$ hold condition for write-after-write access (Fig. 22) forced occupancy of traps in $PU1$ and emptied traps in $PU2$. These traps maintain their occupancy state even though the gate biases are changed after the first write

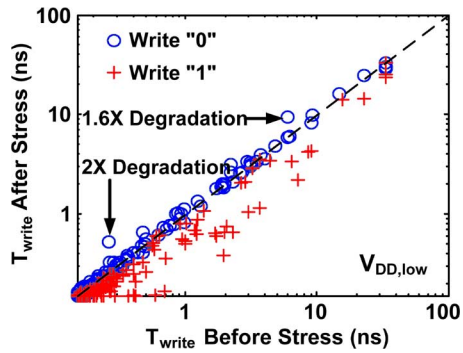


Fig. 24. Data-dependent improvement and degradation in $T_{\rm write}$ due to NBTI stress.

operation because these traps have much longer time constants compared to the 200 ns access time. Trap occupancy in $PU1$ or trap vacancy in $PU2$, set up by the first write operation, degraded writeability of the cell, compared to single write access (Fig. 23(a) and (b)). Read-after-write improved $T_{\rm access}$ of bitcell C compared to single read access (Fig. 23(c)). $T_{\rm access}$ is degraded under single read access because the 100 ms $CL = 0$ hold condition automatically applied a positive gate bias on $PD1$, forcing trap occupancy in $PD1$ which degraded $PD1$, $T_{\rm access}$, and $T_{\rm read}$. These results indicate that dynamic stability should be characterized with write-after-write and single read access in order to capture the worst-case impact of RTN on dynamic stability.

2) Negative-Bias Temperature Instability (NBTI): NBTI refers to degradation in $V_{\rm th}$ of PMOS transistors that is accelerated by negative gate bias and increased temperature. While the impact of NBTI on read stability has been studied extensively, the impact on write stability has mostly been ignored because NBTI actually improves static write margins by degrading $V_{\rm th}$ of the PMOS transistors [22], [23]. Analysis of the sensitivities of $T_{\rm write}$ to transistor variability in Table I leads to the prediction that NBTI actually improves $T_{\rm write}$ of one side and degrades $T_{\rm write}$ of the opposite side of the bitcell, due to the opposite sensitivities of transistors $PU1$ and $PU2$. We experimentally verified this point by subjecting the bitcells to data-dependent NBTI stress while monitoring $T_{\rm write}$ before and after stress. The SRAM array was first initialized to a "0" state. $V_{\rm CELL}$ was then raised to 1.8 V and the test chip was baked at 125 °C for 2 hours. The stored "0" state automatically applied NBTI degradation to only one PMOS transistor in the SRAM bitcell. Since positive-bias temperature instability (PBTI) is not expected in this process technology, only the

transistor characteristics of this one particular PMOS transistor was expected to change from pre-stress to post-stress conditions. Fig. 24 plots measurements of $T_{\rm write}$ before and after stress indicating data-dependent improvement and degradation in write stability. Degradation in $T_{\rm write}$ due to NBTI translates to degradation in maximum frequency $(F_{\rm max})$ of a product or product failure at a given operating frequency.

VI. Conclusion

A dynamic SRAM stability characterization architecture is implemented in 45 nm CMOS. Expected correlations between dynamic stability and static margins were observed in addition to observation of large uncorrelated outliers (10 times more than expected) that are primarily caused by extra PMOS drain resistance. This finding exemplifies the inadequacy of low frequency voltage screen tests for identifying early failures and necessitates at-speed test. $V_{\rm CELL}$ and $V_{\rm WL}$ bias voltages were observed to be effective tuning knobs for balancing critical read stability and writeability but need to be used with caution, due to the enhanced sensitivity of dynamic stability to these biases. Large-amplitude low-frequency RTN signaling in SRAM transistors causes shifts in dynamic stability of similar magnitude that depends on bitcell access patterns. Critical writeability magnifies the impact of process-induced and temporal variability in transistor characteristics, compared to static write margins.

Acknowledgment

The authors wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center, and wafer fabrication donation of STMicroelectronics. They are particularly grateful to Dr. Yasumasa Tsukamoto and Dr. Liang-Teck Pang.

References

[1] International Technology Roadmap for Semiconductors, "International technology roadmap for semiconductors 2009 update system drivers," 2009 [Online]. Available: http://www.itrs.net/Links/2009ITRS/Home2009.htm., [Accessed: Jul. 12, 2010]

[2] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SSC-22, no. 5, pp. 748–754, Oct. 1987.

[3] C. Wann *et al.*, "SRAM cell design for stability methodology," in *VLSI-TSA Dig.*, 2005, pp. 21–22.

[4] R. V. Joshi, S. Mukhopadhyay, D. W. Plass, Y. H. Chan, C.-T. Chuang, and A. Devgan, "Variability analysis for sub-100 nm PD/SOI CMOS SRAM Cell," in *Proc. 30th ESSCC*, 2004, pp. 211–214.

[5] D. Khalil, M. Khellah, N.-S. Kim, Y. Ismail, T. Karnik, and V. K. De, "Accurate estimation of SRAM dynamic stability," *IEEE Trans. VLSI*, vol. 16, no. 12, pp. 1639–1647, Dec. 2008.

[6] M. Sharifkhani and M. Sachdev, "SRAM cell stability: A dynamic perspective," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 609–619, Feb. 2009.

[7] S. O. Toh, Z. Guo, and B. Nikolić, "Dynamic SRAM stability characterization in 45 nm CMOS," in *Symp. VLSI Circuits Dig.*, 2010, pp. 35–36.

[8] Z. Guo, A. Carlson, L.-T. Pang, K. T. Duong, T.-J. K. Liu, and B. Nikolić, "Large-scale SRAM variability characterization in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 11, pp. 3174–3192, Nov. 2009.

[9] A. Bhavnagarwala *et al.*, "Fluctuation limits and scaling opportunities for CMOS SRAM cells," in *IEDM Tech. Dig.*, 2005, pp. 675–678.

[10] A. Carlson, Z. Guo, S. Balasubramanian, L.-T. Pang, T.-J. King, and B. Nikolić, "FinFET SRAM with enhanced read/write margins," in *IEEE Int. SOI Conf.*, 2006, pp. 105–106.

[11] R. Joshi *et al.*, "6.6+ GHz low Vmin, read and half select disturb-free 1.2 Mb SRAM," in *Symp. VLSI Circuits Dig.*, 2007, pp. 250–251.

[12] Y. Morita *et al.*, "Small-defect detection in sub-100 nm SRAM cells using a WL-pulse timing-margin measurement scheme," in *Symp. VLSI Circuits Dig.*, 2010, pp. 37–38.

[13] J. Tsai, S. O. Toh, Z. Guo, L.-T. Pang, T.-J. K. Liu, and B. Nikolić, "SRAM stability characterization using tunable ring oscillators in 45 nm CMOS," in *IEEE ISSCC Dig.*, 2010, pp. 354–355.

[14] L. Wang, N. Xu, S. O. Toh, A. Neureuther, T.-J. K. Liu, and B. Nikolić, "Parameter-specific ring oscillator for process monitoring at the 45 nm node," in *Proc. IEEE CICC*, 2010.

[15] E. Josse *et al.*, "A cost-effective low power platform for the 45-nm technology node," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[16] M. Ball *et al.*, "A screening methodology for VMIN drift in SRAM arrays with application to sub-65 nm nodes," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[17] K. Nii *et al.*, "A 45-nm bulk CMOS embedded SRAM with improved immunity against process and temperature variations," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 180–191, Jan. 2008.

[18] N. Tega *et al.*, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," in *Symp. VLSI Technology Dig.*, 2009, pp. 50–51.

[19] M. Agostinelli *et al.*, "Erratic fluctuations of SRAM cache Vmin at the 90 nm process technology node," in *IEDM Tech. Dig.*, 2005, pp. 655–658.

[20] S. O. Toh, Y. Tsukamoto, Z. Guo, L. Jones, T-J. K. Liu, and B. Nikolić, "Impact of random telegraph signals on Vmin in 45 nm SRAM," in *IEDM Tech. Dig.*, 2009, pp. 767–770.

[21] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise," *Advances in Physics*, pp. 367–468, 1989.

[22] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," in *ISQED Dig.*, 2006, pp. 210–218.

[23] J. C. Lin *et al.*, "Prediction and control of NBTI—Induced SRAM $\rm V_{ccmin}$ drift," in *IEDM Tech. Dig.*, 2006, pp. 1–4.

[24] X. Deng *et al.*, "Characterization of bit transistors in a functional SRAM," in *Symp. VLSI Circuits Dig.*, 2008, pp. 44–45.

[25] J. S. Kolhatkar, "Modeling of RTS noise in MOSFETs under steady-state and large-signal excitation," in *IEDM Tech. Dig.*, 2004, pp. 759–762.

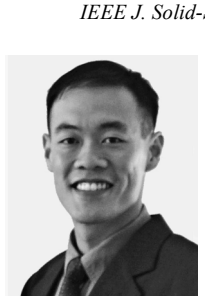
[26] M. Yamaoka, K. Osada, and T. Kawahara, "A cell-activation-time controlled SRAM for low-voltage operation in DVFS SoCs using dynamic stability analysis," in *ESSCIRC Dig.*, 2008, pp. 286–289.

[27] S. O. Toh, T.-J. K. Liu, and B. Nikolić, "Impact of random telegraph signaling noise on SRAM stability," in *Symp. VLSI Technology Dig.*, 2011, pp. 204–205.

[28] K. Zhang *et al.*, "A 3 GHz 70 Mb SRAM in 65 nm CMOS technology with integrated column-based dynamic power supply," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 146–151, Jan. 2006.

[29] M. Yamaoka, N. Maeda, Y. Shimazaki, and K. Osada, "65 nm low-power high-density SRAM operable at 1.0 V under 3σ systematic variation using separate $\rm V_{th}$ monitoring and body bias for NMOS and PMOS," in *IEEE ISSCC Dig.*, 2008, pp. 384–385.

[30] N. Shibata, H. Kiya, S. Kurita, H. Okamoto, M. Tan'no, and T. Douseki, "A 0.5 V 25 MHz 1 mW 256 kb MTCMOS/SOI SRAM for solar-power-operated portable personal digital equipment-sure write operation by using step-down negatively overdriven bit-line scheme," *IEEE J. Solid-State Circuits*, vol. 41, no. 3, pp. 728–742, Mar. 2006.



Seng Oon Toh (S'01) received the B.S. degree (highest honors) in computer engineering from the Georgia Institute of Technology, Atlanta, in 2002. He received the M.S. degree in electrical engineering from the University of California at Berkeley in 2008, where he continues to work towards the Ph.D. degree.

He is also currently working at Advanced Micro Devices on process technology development. His research emphasis is on power-performance optimization as well as robust design of nanoscale SRAM, with emphasis on dynamic stability, RTN, and BTI.

Mr. Toh was awarded an IBM Ph.D. fellowship in 2010 and won the DAC/ISSCC student design contest in 2011.

Zheng Guo (S'03–M'09) received the B.S. degree in computer engineering from the University of Illinois at Urbana-Champaign in 2003. He attended the University of California at Berkeley in the fall of 2003. There, he joined the Berkeley Wireless Research Center in 2004 as a graduate student researcher under the supervision of Professor Borivoje Nikolić. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 2005 and 2009, respectively. Upon completion of his Ph.D. program, he joined Intel in early 2010 as an SRAM technologist.

His Ph.D. research emphasis was on variability characterization and robust design techniques for nanoscale SRAM. In 2004, he was awarded the National Defense Science and Engineering Graduate (NDSEG) Fellowship. In 2005, he received the Best Paper Award at the ACM/IEEE International Symposium of Low-Power Electronics. In 2009, his research work was recognized as a winner in the 46th DAC/ISSCC Student Design Contest.

Tsu-Jae King Liu (F'07) received the B.S., M.S. and Ph.D. degrees in electrical engineering from Stanford University in 1984, 1986 and 1994, respectively.

From 1992 to 1996 she worked as a Member of Research Staff at the Xerox Palo Alto Research Center. In August 1996 she joined the faculty of the University of California at Berkeley, where she is now the Conexant Systems Distinguished Professor of Electrical Engineering and Computer Sciences (EECS) and Associate Dean for Research in the College of Engineering.

Dr. Liu's awards include the Defense Advanced Research Projects Agency Significant Technical Achievement Award (2000) for development of the FinFET, the IEEE Kiyo Tomiyasu Award (2010) for contributions to nanoscale MOS transistors, memory devices, and MEMs devices, and the Electrochemical Society Thomas D. Callinan Award (2011) for excellence in dielectrics and insulation investigations. Her research activities are presently in energy-efficient electronic devices and technology, as well as materials, processes, and devices for integrated microsystems. She has served on committees for various technical conferences, including the IEEE International Electron Devices Meeting and the IEEE Symposium on VLSI Technology, and was an Editor for the IEEE ELECTRON DEVICE LETTERS from 1999 to 2004.

Borivoje Nikolić (S'93–M'99–SM'05) received the Dipl.Ing. and M.Sc. degrees in electrical engineering from the University of Belgrade, Serbia, in 1992 and 1994, respectively, and the Ph.D. degree from the University of California at Davis in 1999.

He lectured electronics courses at the University of Belgrade from 1992 to 1996. He spent two years with Silicon Systems, Inc., Texas Instruments Storage Products Group, San Jose, CA, working on disk-drive signal processing electronics. In 1999, he joined the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, where he is now a Professor. His research activities include digital and analog integrated circuit design and VLSI implementation of communications and signal processing algorithms. He is co-author of the book *Digital Integrated Circuits: A Design Perspective* (Prentice-Hall, 2003).

Dr. Nikolić received the NSF CAREER award in 2003, College of Engineering Best Doctoral Dissertation Prize and Anil K. Jain Prize for the Best Doctoral Dissertation in Electrical and Computer Engineering at University of California at Davis in 1999, as well as the City of Belgrade Award for the Best Diploma Thesis in 1992. For work with his students and colleagues he received the best paper awards at the ISSCC, Symposium on VLSI Circuits, ISLPED and the International SOI Conference.