

Large-Scale SRAM Variability Characterization in 45 nm CMOS

Zheng Guo, *Student Member, IEEE*, Andrew Carlson, *Member, IEEE*, Liang-Teck Pang, *Member, IEEE*, Kenneth T. Duong, *Fellow, IEEE*, and Borivoje Nikolić, *Senior Member, IEEE*

Abstract—Increased process variability presents a major challenge for future SRAM scaling. Fast and accurate validation of SRAM read stability and writeability margins is crucial for estimating yield in large SRAM arrays. Conventional SRAM read/write metrics are characterized through test structures that are able to provide limited hardware measurement data and cannot be used to investigate cell bit fails in functional SRAM arrays. This work presents a method for large-scale characterization of read stability and writeability in functional SRAM arrays using direct bit-line measurements. A test chip is implemented in a 45 nm CMOS process. Large-scale SRAM read/write metrics are measured and compared against conventional SRAM stability metrics. Results show excellent correlation to conventional SRAM read/write metrics as well as V_{MIN} measurements near failure.

Index Terms—CMOS, measurement, noise margins, SRAM, variability.

I. INTRODUCTION

CONTINUED increase in process variability is perceived to be a major challenge to future technology scaling. The impact of process variability is particularly pronounced in large memory arrays due to both the utilization of minimum sized transistors and their extremely large data capacity. To satisfy the functionality of hundreds of millions of SRAM cells in current on-die cache memories, the design has to provide more than 6 standard deviations of margin for parameter variations.

Traditionally, SRAM cell margins have been estimated through SPICE and TCAD simulations, which largely depend on the accuracy of models utilized. However, as processes become increasingly complex and harder to control, designers can no longer rely on model accuracy to fully capture the random effects in large cache memories. Recently, methods have been developed to characterize SRAM variability by measuring DC read/write margins in small SRAM macros with wired-out storage nodes [1], [2]. This significantly enhances the accuracy

Manuscript received April 30, 2009; revised August 28, 2009. Current version published October 23, 2009. This paper was approved by Associate Editor Peter Gillingham. This work was supported by the National Science Foundation Infrastructure Grant No. 0403427, a chip fabrication donation of STMicroelectronics, the Center for Circuit & System Solutions (C2S2) Focus Center, and the IBM Faculty Partnership Award.

Z. Guo, T.-J. K. Liu, and B. Nikolić are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: zhengguo@eecs.berkeley.edu; bora@eecs.berkeley.edu).

A. Carlson is with Advanced Micro Devices.

L.-T. Pang is with IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

K. T. Duong is with Sun Microsystems.

Digital Object Identifier 10.1109/JSSC.2009.2032698

of SRAM failure analysis over TCAD simulations but requires the removal of upper metal layers and the insertion of large switch networks to access all internal storage nodes. As a result, this approach is limited to smaller data volumes that may be unsuitable for failure analysis of large cache memory. Thus, SRAM designers continue to rely on collecting distributions of bit-line read currents [4] and minimum operating voltage (V_{MIN}) [5], [6] to gauge SRAM read stability and writeability in large functional SRAM arrays. However, direct correlation between measured SRAM read/write margins and V_{MIN} in large functional SRAM arrays has not been established.

In this work, we propose a method to characterize SRAM cell read stability and writeability margins through direct bit-line measurements in large functional SRAM arrays [3]. The characterization of SRAM read stability and writeability in functional cache memory not only provides substantial hardware data for statistical and failure analysis but can also complement standard SRAM built-in self test (BIST) methods by correlating BIST failures to the measured bit cell read/write margins. Furthermore, irregular bit cell characteristics measured through direct bit-line access can be mapped to the cell location and verified using nano-probing to determine its source. The proposed method is extended to characterize V_{MIN} for each SRAM cell in the array during standby, read, and write cycles. As a result, a direct correlation between measured SRAM read/write margins and V_{MIN} in functional SRAM arrays is established in this work. This characterization methodology is validated in a commercial low-power 45 nm CMOS process. The test chip also includes small SRAM macros with all-internal-node access, similar to those in [1], [2], to correlate large-scale and macro DC read/write margin measurements.

We first review the conventional SRAM stability metrics in Section II. Section III introduces the large-scale read stability and writeability metrics and describes how direct bit-line measurements can be applied to V_{MIN} characterization. The design implementation is detailed in Section IV. Section V presents the 45 nm CMOS test chip. Section VI summarizes the measurement results and their implications. Finally, the conclusions are drawn in Section VII.

II. CONVENTIONAL SRAM DESIGN METRICS

A. Read Stability

1) *Read Static Noise Margin (RSNM)*: Fig. 1(a) shows a schematic of a 6-T SRAM cell. It consists of two cross coupled inverters ($P_L - N_L$ and $P_R - N_R$) for data retention and two pass-gates (N_{AXL} and N_{AXR}) for read/write access. During a

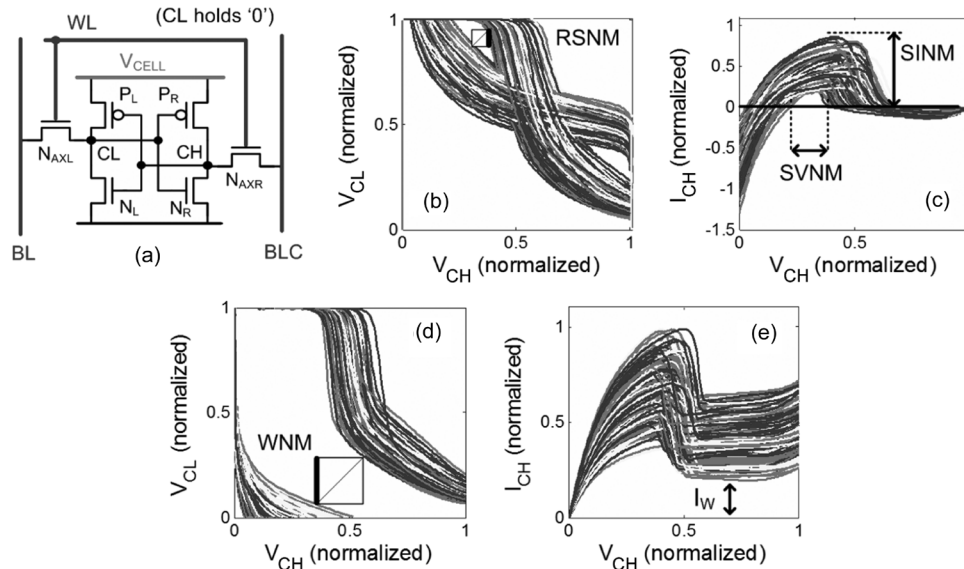


Fig. 1. (a) Schematic of a 6-T SRAM cell. Definition of (b) RSNM from measured VTC, (c) SVN and SINM from measured N-curve. Definition of (d) WNM from measured VTC and (e) I_W from measured N-curve.

read operation, storage node voltage V_{CL} rises above 0 V, to a voltage determined by the resistive voltage divider set up by the pass-gate N_{AXL} and the pull-down N_L between BL and storage node CL. If V_{CL} exceeds the trip point of inverter $P_R - N_R$ during the read cycle, the cell bit will flip, causing a read upset. The most common metric used for SRAM read stability is the read static noise margin (RSNM), which represents the maximum tolerable DC noise voltage at each storage node before causing a read upset and is extracted from the read voltage transfer characteristics (VTC) [7]. The read VTC can be measured by sweeping the voltage at the storage node CH (or CL) with both bit-lines (BL, BLC) and word-line (WL) biased at V_{DD} while monitoring the node voltage at CL (or CH). The RSNM can be quantified by the side of the largest square embedded between the read VTC measured from the same SRAM cell [Fig. 1(b)].

2) *Static Voltage Noise Margin (SVNM) and Static Current Noise Margin (SINM)*: A second metric used to characterize SRAM read stability makes use of the N-curve [8], [9] measured by sweeping the voltage at the storage node CH (or CL) with BL, BLC, and WL biased at V_{DD} while monitoring the current externally sourced into the CH (or CL) node. The N-curve provides both voltage and current information in characterizing the read stability of an SRAM cell. The three points crossing the x axis (where $I_{CH} = 0$) on the N-curve [Fig. 1(c)] correspond to the three intersection points between the two sets of read VTC [Fig. 1(b)]. The static voltage noise margin (SVNM) [9] of an SRAM cell can be quantified by the voltage difference between the first two points crossing the x axis as the voltage at the storage node CH (or CL) is ramped up [Fig. 1(c)]. The peak current between these two points, effectively measuring the maximum tolerable DC current injected into the storage node of an SRAM cell without disturbing its data, quantifies its static current noise margin (SINM) [9].

B. Writeability

1) *Write Noise Margin (WNM)*: During a write operation, N_{AXR} and P_R form a resistive voltage divider for the falling BLC and the storage node CH [Fig. 1(a)]. If the voltage divider pulls V_{CH} below the trip point of inverter $P_L - N_L$, a successful write operation occurs. SRAM writeability can be gauged by the write noise margin (WNM) [1], [2] extracted using a combination of the read VTC and the write VTC. The write VTC is measured by sweeping the voltage at the storage node CL [Fig. 1(d), y-axis] with BL and WL biased at V_{DD} and BLC biased at V_{SS} while monitoring the node voltage at CH [Fig. 1(d), x-axis].¹ This write VTC should be used in combination with the read VTC measured by sweeping the voltage at the storage node CH [Fig. 1(d), x-axis] while monitoring the node voltage at CL [Fig. 1(d), y-axis]. The WNM can be quantified by the side of the smallest square embedded between the read and the write VTC measured from the same SRAM cell [Fig. 1(d)] at the lower half of the curves, past the inverter $P_L - N_L$ trip point. When WNM falls below zero, the write VTC intersects the read VTC, indicating a positive retention margin even when BLC is tied to V_{SS} , thus suggesting an inability to write.

2) *Writeability Current (I_W)*: SRAM writeability can also be characterized using the N-curve. Unlike the N-curve setup to characterize read stability, the N-curve for writeability characterization is measured by sweeping the voltage at the storage node CH (CL) with BL (BLC) and WL biased at V_{DD} and BLC (BL) biased at V_{SS} while monitoring the current externally sourced into the CH (CL) node. This effectively measures N_{AXR} current minus P_R current. The writeability current [10], [11], I_W , is defined as the minimum measured current past the inverter $P_L - N_L$ trip point [Fig. 1(e)]. A larger I_W corresponds

¹Write VTC for writing a '0' to CL can be measured by sweeping V_{CH} with BLC and WL at V_{DD} and BL at V_{SS} while monitoring V_{CL} . This write VTC should be used in combination with the read VTC measured by sweeping V_{CL} .

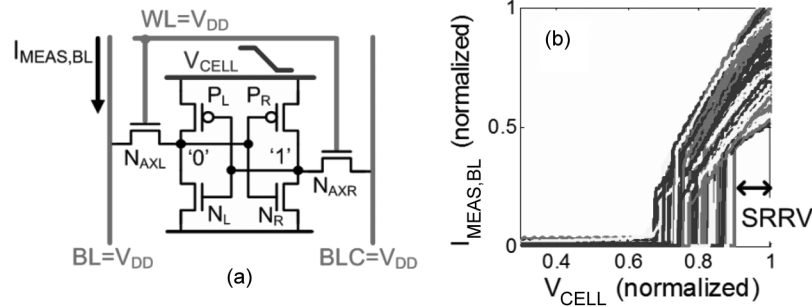


Fig. 2. (a) Schematic setup for supply read retention voltage (SRRV) measurements. (b) Definition of SRRV from measured transfer curves.

to a more writeable cell, while $I_W < 0$ represents a write failure. It should be noted that if $I_W \leq 0$, the write VTC will intersect the read VTC [in Fig. 1(d)] at the same V_{CH} point(s) where the measured current I_{CH} intersects the x axis [in Fig. 1(e)], resulting in zero or negative WNM.

III. LARGE-SCALE SRAM DESIGN METRICS

The major drawback of the conventional DC read/write margin metrics presented in the previous section is the inability to measure them in dense functional SRAM arrays because of the metal spacing constraints for routing out internal storage nodes and the significant area overhead associated with the switch array. This results in an insufficient number of data points for failure analysis of large cache memories. To increase the sample size, the SRAM array must stay intact; in this case, SRAM read stability and writeability must be characterized by accessing only the bit-lines, the word-line, and the cell supply voltages. As an example, bit-line access has been previously applied to detect and isolate faulty SRAM cells in memory arrays [12]. Similarly, large-scale performance of the SRAM cells has been characterized through distributions of per-cell minimum operating voltages [5], [6]. In this section, we introduce a method for characterizing the SRAM cell read stability and writeability in functional SRAM arrays by taking advantage of direct bit-line measurements while adjusting bit-line, word-line, and cell supply voltages. Furthermore, we show how direct bit-line measurements can be adopted to characterize the SRAM V_{MIN} during standby, read, and write cycles.

A. Read Stability Measurement

1) *Supply Read Retention Voltage (SRRV)*: During the read cycle, both bit-lines float around V_{DD} while the word-line is driven high, and the cell state is retained by keeping the cell supply sufficiently high. The SRAM read stability in functional SRAM arrays can be gauged by the lowest cell supply voltage for data retention during a read cycle, denoted as the supply read retention voltage (SRRV) [3]. Fig. 2(a) shows the measurement setup for SRRV. To capture the SRAM read stability, both BL and BLC are precharged and WL is driven by the operating voltage (V_{DD}) to emulate a read cycle. The BL current at the '0' storage node ($I_{MEAS,BL}$) is monitored while ramping down the SRAM cell supply (V_{CELL}). When the cell supply is dropped sufficiently low, the SRAM cell loses its ability for

data retention when N_{AXL} dominates N_L so that CL, originally holding '0', rises above the trip point of inverter $P_R - N_R$. At that point, the cell state flips, signified by a sudden drop in $I_{MEAS,BL}$. The measured transfer curves, of $I_{MEAS,BL}$ versus V_{CELL} , are plotted in Fig. 2(b). The difference between V_{DD} and the value of V_{CELL} causing $I_{MEAS,BL}$ to suddenly drop represents the SRRV of the SRAM cell. When $SRRV = 0$, the SRAM cell is biased for a nominal read operation with WL, BL, BLC and V_{CELL} all biased at V_{DD} . $SRRV > 0$ indicates that V_{CELL} can be dropped below V_{DD} without disturbing the data. Therefore, SRRV represents the maximum tolerable DC noise voltage at the cell supply before causing a read upset.

Intrinsic mismatch of transistors within an SRAM cell typically results in an asymmetry in the cell robustness to read upset between storing a '1' (when CL holds '0') and storing a '0' (when CH holds '0'). Depending on the degree of asymmetry in the SRAM cell, a data disturbance, in the form of a bit flip, either occurs on both sides of the SRAM cell or only on the less read-stable side as the cell supply is dropped. Fig. 3(a) and (b) shows the SRRV transfer curves for SRAM cells with worse read stability when CL holds '0'—i.e., lower cell β -ratio (defined as the strength ratio of pull-down to pass-gate transistors) at the CL node. Fig. 3(a) shows that when '0' is stored at the less read-stable CL node, all curves exhibit a sharp fall-off in the BL current ($I_{MEAS,BL}$), suggesting a clear SRAM cell data disturbance in the form of a bit flip. However, when '0' is stored at the more read-stable CH node, only some of the curves exhibit a sharp fall-off in the BLC current ($I_{MEAS,BLC}$), corresponding to SRAM cells with a smaller asymmetry between the two halves, while other curves show a smooth bending in the measured $I_{MEAS,BLC}$ [Fig. 3(b)], corresponding to SRAM cells with a larger asymmetry between the two halves. In the latter case, due to a heavily skewed read stability favoring the node CH, a clear data disturbance, in the form of a bit flip, does not occur when the cell supply is dropped beyond data retention and the SRAM cell enters a metastable state. Fig. 3(c) and (d) shows similar SRRV transfer curves for SRAM cells with worse read stability when CH holds '0'—i.e., lower cell β -ratio at the CH node. To gauge SRAM read stability, the SRRV value extracted from the less read-stable storage node is used.

2) *Word-Line Read Retention Voltage (WRRV)*: When the word-line is driven high during a read/write cycle, both the SRAM cell under direct read access and all unaccessed SRAM

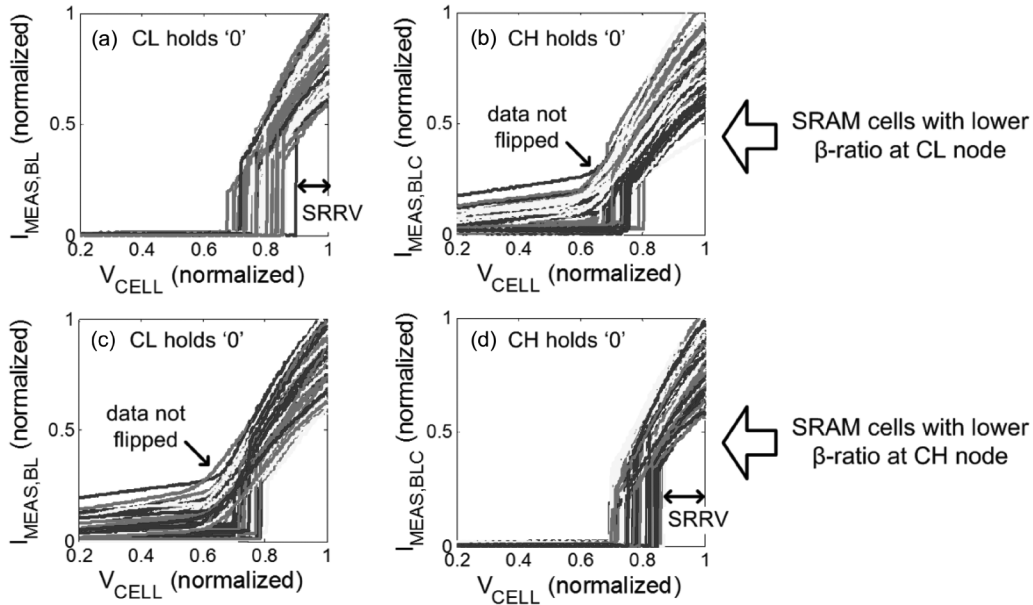


Fig. 3. (a) SRRV transfer curves for storing ‘0’ at the less read-stable CL node; all curves exhibit sharp fall off in current, indicating data disturbance. (b) SRRV transfer curves for storing ‘0’ at the more read-stable CH node; some curves exhibit smooth bending, indicating bits do not flip. (c) SRRV transfer curves for storing ‘0’ at the more read-stable CL node; some curves exhibit smooth bending, indicating bits do not flip. (d) SRRV transfer curves for storing ‘0’ at the less read-stable CH node; all curves exhibit sharp fall off in current, indicating data disturbance.

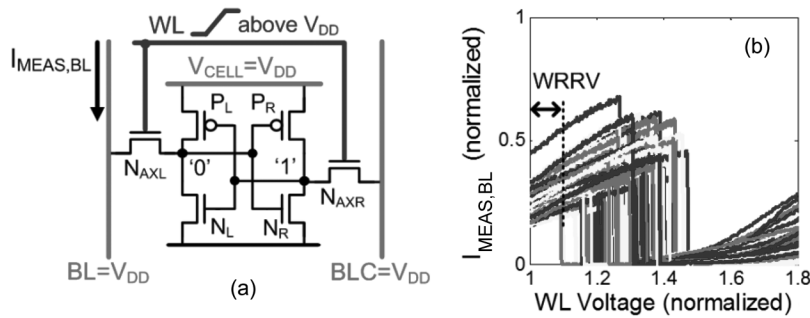


Fig. 4. (a) Schematic setup for word line read retention voltage (WRRV) measurements. (b) Definition of WRRV from measured transfer curves.

cells driven by the asserted word-line undergo a read stress. This read stress can be exacerbated by boosting the word-line voltage beyond V_{DD} . Therefore, the read stability of an SRAM cell can also be gauged by the largest word-line boost without upsetting cell data retention, denoted as the word-line read retention voltage (WRRV). Fig. 4(a) shows the measurement setup for WRRV. To capture the WRRV of the SRAM cell, the SRAM cell supply (V_{CELL}) is biased at V_{DD} with both BL and BLC precharged. WL voltage is ramped above V_{DD} , and kept below the gate-oxide breakdown voltage set by the technology, while the BL current at the ‘0’ storage node ($I_{MEAS,BL}$) is monitored. When the WL voltage is boosted sufficiently high above V_{DD} , the SRAM cell state is disturbed due to an exacerbated read stress as N_{AXL} dominates N_L and pulls V_{CL} above the trip point of inverter $P_R - N_R$. The cell disturbance is captured as a sudden drop in the measured current $I_{MEAS,BL}$. The measured transfer curves, of $I_{MEAS,BL}$ versus WL voltage, are plotted in Fig. 4(b). The WRRV of an SRAM cell is quantified as the difference between the WL voltage causing $I_{MEAS,BL}$ to suddenly drop and V_{DD} . Similar to SRRV, when $WRRV = 0$,

the SRAM cell is biased for a nominal read operation with WL, BL, BLC and V_{CELL} all biased at V_{DD} . $WRRV > 0$ indicates that V_{WL} can be boosted above V_{DD} without disturbing the data. Therefore, WRRV represents the maximum tolerable DC voltage rise on the WL before causing a read upset.

When the read stability of the SRAM cell becomes heavily skewed towards either storage node CH or CL, data stored in the more robust node will be preserved even under very high WL boost, and the measured current $I_{MEAS,BL}$ never drops significantly. Fig. 5(a) and (b) plots the WRRV transfer curves for SRAM cells with worse read stability when CL holds ‘0’—i.e., lower cell β -ratio at CL node. Fig. 5(a) shows that when ‘0’ is stored at the less read-stable CL node, all curves exhibit a sharp drop in $I_{MEAS,BL}$, suggesting a clear SRAM cell data disturbance. However, when ‘0’ is stored at the more read-stable CH node, only some curves exhibit a sharp drop in $I_{MEAS,BLC}$, corresponding to SRAM cells with a smaller asymmetry between the two halves, while other curves show either continued increase or a slight bending in the measured $I_{MEAS,BLC}$ [Fig. 5(b)], corresponding to SRAM cells with a

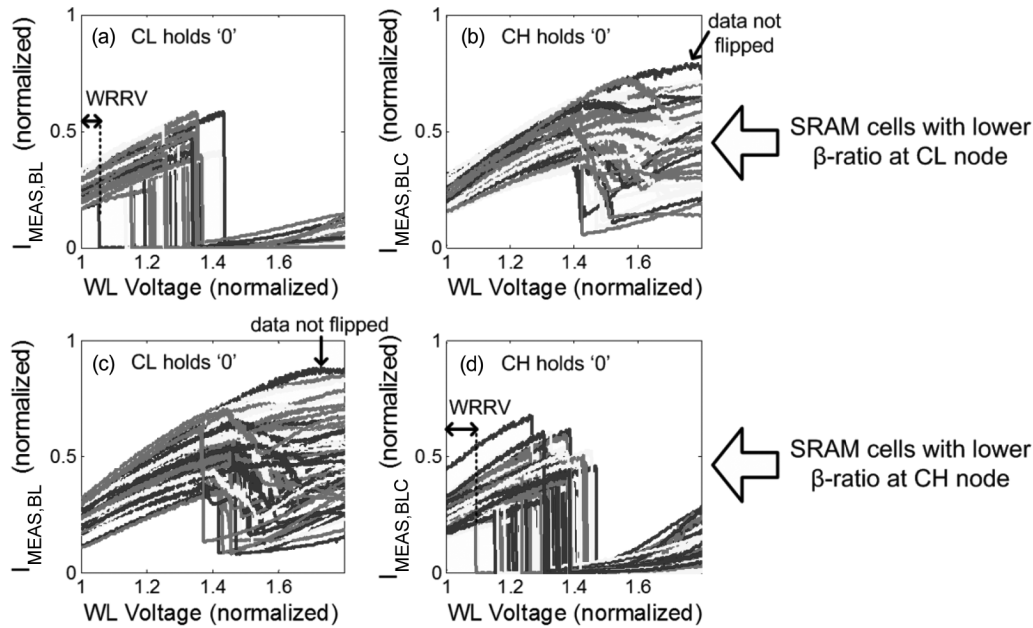


Fig. 5. (a) WRRV transfer curves for storing '0' at the less read-stable CL node; all curves exhibit sharp drop in current, indicating data disturbance. (b) WRRV transfer curves for storing '0' at the more read-stable CH node; some curves continue to increase or bend slightly, indicating bits do not flip. (c) WRRV transfer curves for storing '0' at the more read-stable CL node; some curves continue to increase or bend slightly, indicating bits do not flip. (d) WRRV transfer curves for storing '0' at the less read-stable CH node; all curves exhibit sharp drop in current, indicating data disturbance.

larger asymmetry between the two halves. In the latter case, due to a heavily skewed read stability favoring the node CH, the cell state is not disturbed by the overdriven WL. Fig. 5(c) and (d) plots similar WRRV transfer curves for SRAM cells with worse read stability when CH holds '0'—i.e., lower cell β -ratio at CH node. To gauge SRAM read stability, the WRRV value extracted from the less read-stable storage node is used.

B. Writeability Measurement

1) *Bit-Line Write Trip Voltage (BWTV)*: During the write cycle, bit-lines are configured according to the new data and the word-line is driven high. The writeability of an SRAM cell in a functional SRAM array can be gauged by the maximum bit-line voltage, at the '1' storage node, able to flip the cell state during a write cycle [3], [4], [13], [14], denoted as the bit-line write trip voltage (BWTV). Fig. 6(a) shows the measurement setup for BWTV. To capture the BWTV of an SRAM cell, the cell supply (V_{CELL}), WL, and BL (at the '0' storage node) are biased at V_{DD} . BL current at the '0' storage node ($I_{MEAS,BL}$) is monitored while ramping down the BLC voltage. As BLC is ramped low, the pass-gate N_{AXR} overcomes P_R and the '1' storage is dropped below the inverter P_L-N_L trip point, resulting in a successful write, signified by a sudden drop in $I_{MEAS,BL}$. Fig. 6(b) plots the measured transfer curves of $I_{MEAS,BL}$ versus BLC voltage. The BWTV is quantified as the BLC voltage that induces a sudden change in $I_{MEAS,BL}$. When $BWTV = 0$, the SRAM cell is biased for a nominal write operation with WL, BL (or BLC), and V_{CELL} biased at V_{DD} and BLC (or BL) biased at V_{SS} . $BWTV > 0$ indicates that a successful write operation can take place even with a BLC (or BL) voltage higher than V_{SS} . Therefore, BWTV represents the maximum bit-line voltage slack that can be tolerated to successfully write the cell.

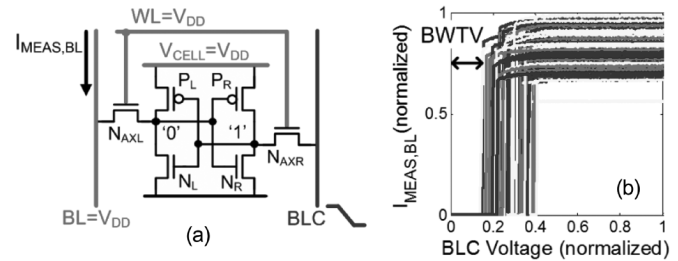


Fig. 6. (a) Schematic setup for bit line write trip voltage (BWTV) measurements. (b) Definition of BWTV from measured transfer curves.

Since the measurement setup for BWTV requires first exerting a read stress to the SRAM cell under test (CUT), a read disturb may occur in the CUT before the BWTV can be observed when testing at lower supply voltages. Due to intrinsic mismatch of transistors within an SRAM cell, a read disturbance at a high enough supply voltage typically happens to only one side of the CUT while a read disturbance at the other side either may happen at a lower supply voltage or not at all (Fig. 3). In this case, BWTV can still be characterized for the more read-stable side of the CUT, which typically corresponds to the less writeable side. Fig. 7 plots the measured transfer curves with $V_{DD} = 0.4$ V for SRAM cells undergoing read disturbance. SRAM cells undergoing read disturbance on both sides were not identified with V_{DD} as low as 0.4 V for the particular chip tested. Over $10\times$ variation in the measured BL current before the stored data of the CUT flips was observed, indicating transistor operation in the onset of weak-inversion. Further reduction of the supply voltage may result in SRAM cells undergoing read disturbance on both sides, in which case the BWTV can no longer be characterized for either side.

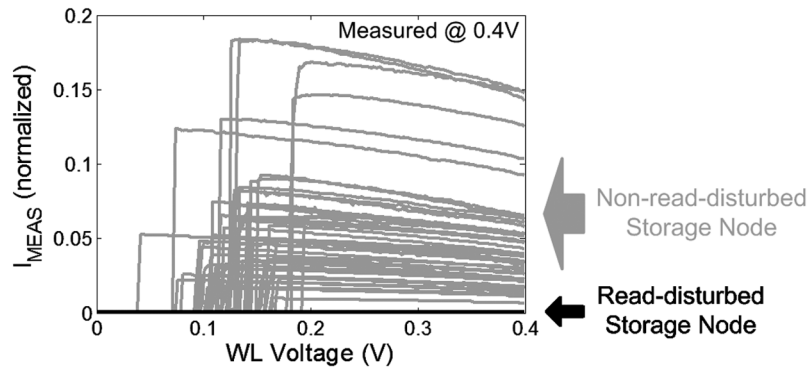


Fig. 7. Measured transfer curves for BWTV characterization at 0.4 V. Curves are plotted only for SRAM cells undergoing a read disturbance. SRAM cells undergoing read disturbance on both sides were not identified at 0.4 V for this particular chip.

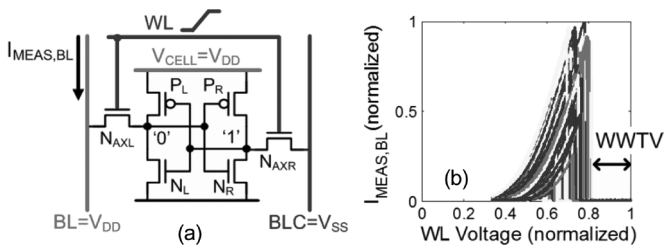


Fig. 8. (a) Schematic setup for word line write trip voltage (WWTV) measurements. (b) Definition of WWTV from measured transfer curves.

2) *Word-Line Write Trip Voltage (WWTV)*: The writeability of an SRAM cell can also be captured by first configuring the bit-lines to write the data and then ramping up the word-line [3], [14]. The minimum word-line voltage able to flip the cell state during a write cycle, denoted as the word-line write trip voltage (WWTV), can be used to gauge the SRAM writeability. Fig. 8(a) presents the measurement setup for WWTV. The cell supply (V_{CELL}) and BL are biased at V_{DD} while BLC is biased at V_{SS} . BL current at the ‘0’ storage node ($I_{MEAS,BL}$) is monitored while ramping up the WL voltage. As V_{WL} is ramped high, the measured current initially resembles the $I_D - V_G$ curve of the pass-gate N_{AXL} . When the WL voltage is sufficiently high, the cell state flips, signified by a sudden drop in the magnitude of $I_{MEAS,BL}$. Fig. 8(b) plots the measured transfer curves of $I_{MEAS,BL}$ versus WL voltage. The WWTV is quantified as the value $(V_{DD} - V_{WL})$, where V_{WL} is the minimum WL voltage causing the sudden drop in $I_{MEAS,BL}$. Similar to BWTV, when $WWTV = 0$, the SRAM cell is biased for a nominal write operation with WL, BL (or BLC), and V_{CELL} biased at V_{DD} and BLC (or BL) biased at V_{SS} . $WWTV > 0$ indicates that a successful write operation can take place even with a WL voltage lower than V_{DD} . Therefore, WWTV represents the maximum WL voltage slack that can be tolerated to successfully write the cell. The most notable advantage of the WWTV measurement is that, unlike during the BWTV measurement, the SRAM CUT is not put under read stress at the onset of the measurement (with $V_{WL} = 0$ V). Therefore, WWTV can continue to be characterized for SRAM cells under aggressively scaled supply voltages.

C. V_{MIN} Characterization

In addition to read stability and writeability characterization, the direct bit-line access scheme can be adopted to characterize the minimum DC operating voltage of each SRAM cell during standby, read, and write cycles. Fig. 9(a) shows the flow-chart diagram for measuring SRAM V_{MIN} during a static read operation. Each iteration of this measurement starts with a data initialization under the nominal supply voltage ($V_{DD,NOMINAL}$)—1.1 V in this technology. The SRAM cell is then configured for a low voltage read operation with V_{CELL} , V_{BL} , and V_{WL} all set to a lower supply voltage, which is gradually reduced for each iteration of the measurement process. Finally, the supply voltage is raised for a read operation at $V_{DD,NOMINAL}$ and the BL current at the ‘0’-initialized storage node ($I_{MEAS,BL}$) is measured. The measured current should be high—equal to I_{READ} —if no data disturbance took place during the low voltage read. Therefore, V_{MIN} can be characterized as the maximum operation voltage before $I_{MEAS,BL}$ drops [Fig. 9(b)]. SRAM V_{MIN} during standby can be characterized using the same procedure as in Fig. 9(a) by keeping V_{WL} at V_{SS} during each low voltage read operation, to emulate a low voltage hold operation. To eliminate accidental data disturbance, WL is turned off between low voltage and high voltage operations. Furthermore, V_{WL} is reduced by 200 mV during the high voltage read operation.

Fig. 9(c) shows the flow chart diagram for measuring SRAM V_{MIN} during a static write operation. The procedure is very similar to that for the read V_{MIN} characterization—but instead of a low voltage read operation, each iteration performs a low voltage write operation where V_{CELL} , V_{BL} at the ‘0’-initialized storage node, and V_{WL} are set to a lower supply voltage and V_{BL} at the ‘1’-initialized storage node is set at V_{SS} . Each low voltage write is immediately followed by a high voltage read where the BL current at the ‘0’-initialized storage node ($I_{MEAS,BL}$) is measured. The measured current should be low if data is successfully written during the low voltage write. V_{MIN} can be characterized as the maximum operation voltage while $I_{MEAS,BL}$ remains low [Fig. 9(d)]. Note that V_{MIN} characterization using direct bit-line measurements is slower than the typical on-chip digital SRAM tester, using similar read-after-read and read-after-write sequences described above, because

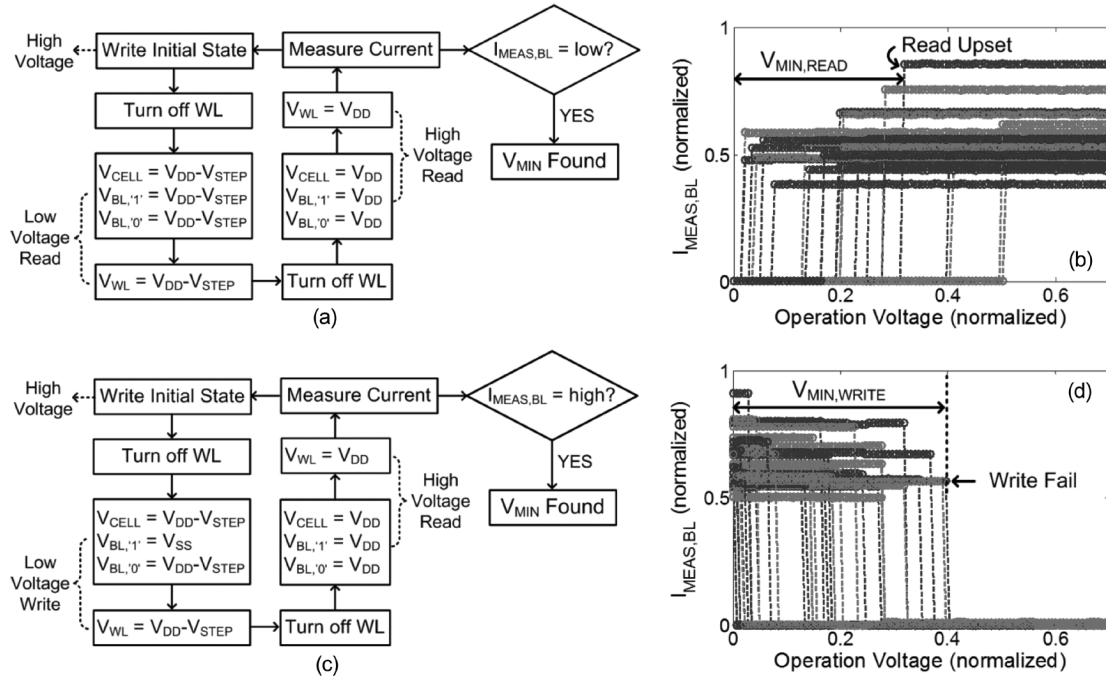


Fig. 9. (a) Flow chart for read V_{MIN} measurement and (b) measured bit line currents at different operation voltages for read V_{MIN} characterization. (c) Flow chart for write V_{MIN} measurement and (d) measured bit line currents at different operation voltages for write V_{MIN} characterization.

of the need to monitor the BL current. However, since the direct bit-line V_{MIN} characterization can be performed alongside the large-scale read stability and writeability measurements with no additional hardware overhead, it is used, in this work, to establish correlations between cell failure and the cell read and write characteristics.

IV. CIRCUIT IMPLEMENTATION

Fig. 10(a) presents the circuit diagram for the direct bit-line characterization scheme for functional SRAM arrays. The lower right portion of the circuit diagram shows a typical functional SRAM array with row decoder and column read/write circuitry. A level shifter with a supply voltage slightly below 0 V ($V_{SS,NEG}$) is inserted in the row decoder to allow a sufficient range of WL voltages from 0 V to +400 mV above $V_{DD,NOMINAL}$ (set by $V_{DD,WL}$). The SRAM array is implemented with independent cell supply (V_{CELL}), cell ground (V_{SSCL}), N-well bias (V_{NW}), and P-well bias (V_{PW}). All four terminals can be used either for voltage sweeping or for setting bias conditions. During direct bit-line measurements, the column read/write circuitry can be shut off by a low R/W enable signal. The switch network for direct bit-line access is implemented using wide, long-channel, thick-oxide CMOS transmission gates driven by a separate supply to suppress leakage in unaccessed bit-lines. The bit-lines are accessed through four levels of hierarchy with a maximum of 16 switches sharing the same node. Leakage in the switch network is minimized by the stack effect in the hierarchy. The only significant source of leakage in the switch network comes from the drain to body leakage, which sets the lower limit of measurable current at a few to a few tens of nA and does not affect

read/write margin measurements. In order to accurately set voltages at the bit-lines, the V_{DS} drop in the switch hierarchy must be eliminated. Overdriving the gates of the thick oxide transistors can help decrease the V_{DS} drop in each switch, but cannot completely mitigate its effects. Thus, we adopted the 4-terminal Kelvin sensing method using independent force (current) and sense (voltage) paths to access each bit-line. This effectively eliminates the V_{DS} drop (series resistance) in the switch hierarchy. All static control signals are supplied by the scan chain to minimize I/O pin count.

The overall area overhead of the bit-line switch network in this prototype is approximately 20%² and can be reduced with an optimized layout of the bit-line switch network and/or by reducing the depth of the switch hierarchy. In addition, the array efficiency is enhanced in SRAM arrays with larger column heights. The proposed direct bit-line characterization scheme requires that the worst-case on-current of a single pass-gate transistor connected to a bit-line be higher than the sum of leakage currents of all pass-gate transistors connected to the complementary bit-line. This requirement is typically less stringent than the constraint set by the SRAM read access performance and therefore should not limit the column segmentation of the SRAM array. However, in the case where the SRAM read access constraint is relaxed and the bit-line leakage is high, direct bit-line measurements at lower operating voltages may be challenging as the bit-line on- and off- currents become harder to distinguish when detecting a data flip. This can be solved by returning to a higher operating voltage that ensures read stability for bit-line current measurements (i.e., a

²The 6-bit DAC is not included in this estimation as it is not required for direct bit-line characterization (see Section V).

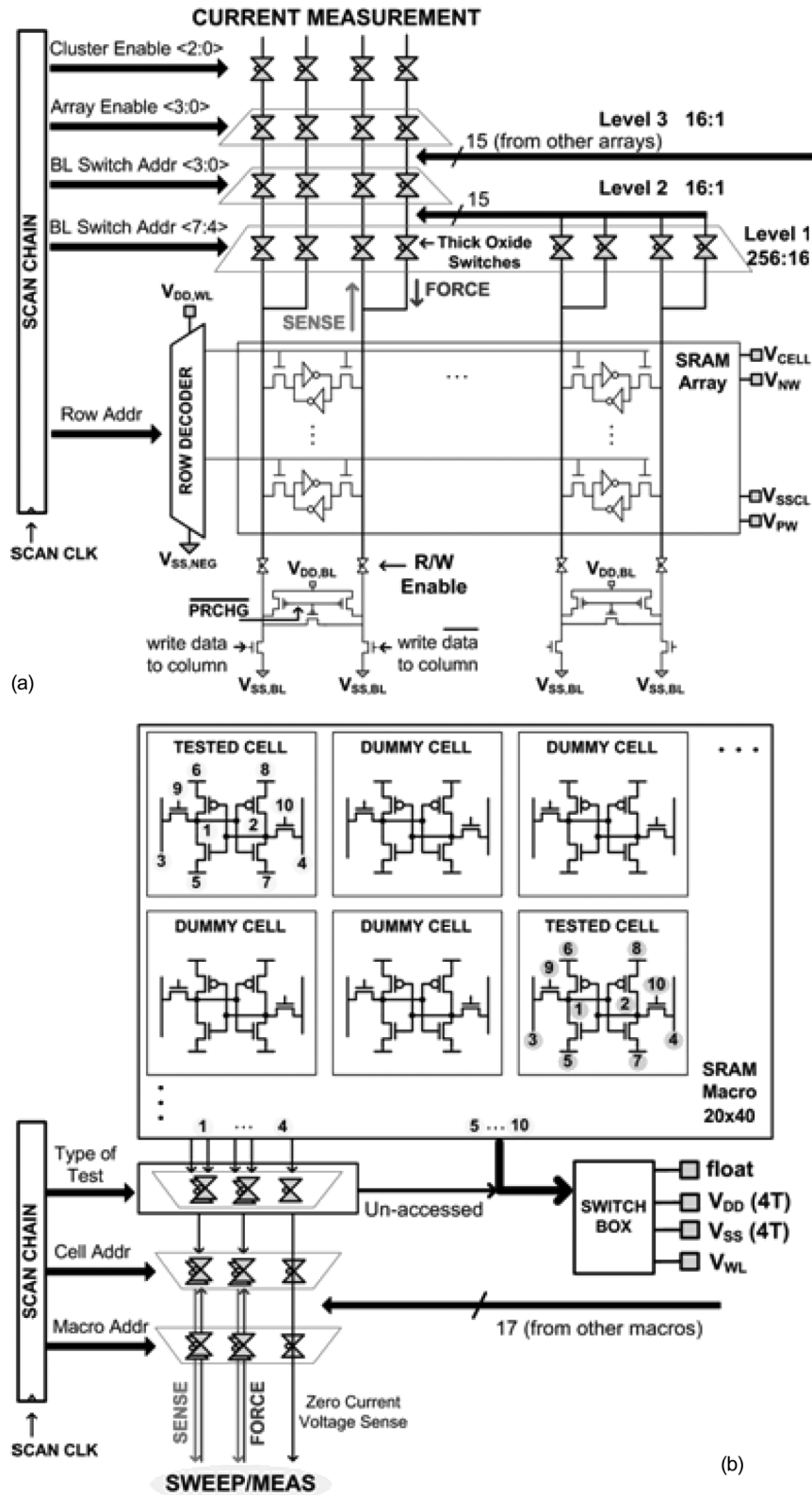


Fig. 10. Circuit diagram of (a) direct bit line measurement scheme in functional SRAM and (b) all-internal-node access scheme in SRAM macros.

high voltage read operation) after stressing the SRAM cell with the appropriate sweeping voltage at a lower supply—similar to the V_{MIN} characterization loop described in Section III-C. Due to a reasonably low overhead, the proposed direct bit-line characterization can either be implemented in an early SRAM devel-

opment vehicle or, occasionally, on a working chip to monitor the process variability.

Fig. 10(b) shows the circuit diagram of the all-internal-node access characterization scheme for small SRAM macros. Each SRAM macro consists of a 20-row by 40-column array, with

one cell accessed per column and per row. All 10 internal nodes— V_{DD} , V_{SS} , V_{WL} , V_{BL} , and the storage node of both sides—of each SRAM cell under test (CUT) are wired out through a hierarchy of switches to allow SRAM VTC curve and N-curve measurements as well as individual transistor I - V characterization. N-well and P-well biasing— V_{NW} and V_{PW} —in each macro is shared with the functional SRAM arrays to investigate the effect of body biasing on SRAM read/write margins and transistor V_{TH} . To provide enough spacing for routing 10 internal nodes out per cell, every other column in the array is skipped, yielding 20 accessed SRAM cells per macro. The 4-terminal Kelvin sensing method is adopted to access the storage nodes and both bit-lines to mitigate the effect of V_{DS} drop in the switch hierarchy—one of the two storage nodes is selectively wired out for zero current voltage sensing, which does not require the 4-terminal Kelvin setup. The other 6 internal nodes from each SRAM cell are selectively left floating or connected to V_{DD} , V_{SS} , or V_{WL} for bias.³ Each internal node is accessed through 3 levels of switching. All static control signals are supplied by the scan chain.

V. 45 NM CMOS TEST CHIP

A 2.2 mm \times 2.2 mm test chip [3], [15], [16] (Fig. 11) is implemented in a low-power strained-Si 45 nm CMOS process [17]–[19] with 7 metal layers. The SRAM cell provided with this process is high-speed (i.e., high read current), with a cell area of $0.374 \mu\text{m}^2$. The test chip consists of eight 32 kb ($128 \text{ WL} \times 256 \text{ BL}$) and eight 64 kb ($256 \text{ WL} \times 256 \text{ BL}$) functional SRAM arrays (768 kb in total) for large-scale read stability, writeability, V_{MIN} , and cell read current (I_{READ}) characterization. It also includes eighteen 20×40 small SRAM macros with all-internal-node access for conventional SRAM VTC, N-curve, and individual transistor I - V measurements. Twenty cells in each SRAM macro have all internal nodes externally accessible through a switch network. A 6-bit DAC is shared by every two 256×256 SRAM arrays for optional on-chip WL sweep but was not used during testing as it does not enhance measurement speed due to equipment limitations.

VI. MEASUREMENT RESULTS

A. Read/Write Margin Metric Comparison

1) μ/σ Comparison: Fig. 12 plots the distribution densities of both SRAM read stability and writeability metrics measured at $V_{DD} = 0.7 \text{ V}$ on a log scale and fitted to Gaussian distributions—using RSNM, SRRV, and WRRV in (a) and I_{W} , BWTV, and WWTV in (b). Each metric is normalized to its σ value. RSNM and I_{W} are measured for several hundred SRAM cells via all-internal-node access in the SRAM macros. SRRV, WRRV, BWTV, and WWTV are measured for hundreds of thousands of SRAM cells in the functional SRAM arrays using direct bit-line access, and so they capture several orders of magnitude more statistical data than RSNM and I_{W} .

³4-terminal Kelvin setup is also applied for V_{DD} and V_{SS} nodes when needed.

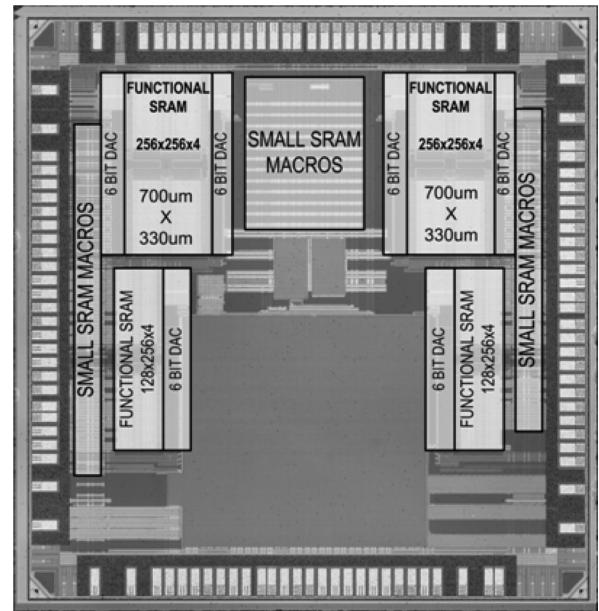


Fig. 11. Die photo of the 45 nm CMOS test chip.

Fig. 12 shows that both read and write metrics, at 0.7 V, exhibit a slightly different μ/σ value with respect to one another. This difference is small among the read metrics and more pronounced among the write metrics and it can be attributed to a few different mechanisms. Fig. 13(a) and (b) plots the μ and σ of the measured SRRV and WRRV as a function of the supply voltage. As the WL is driven above the SRAM cell supply voltage (V_{CELL}), the gate overdrive of the pass-gate transistor at the '0' storage node saturates as the '0' storage node is pulled above V_{SS} . This is manifested in a reduced sensitivity in the measured BL current to the WL voltage—decreasing slope in Fig. 4(b)—as the WL is driven above V_{CELL} . This sensitivity is further reduced at higher supply voltages, as the inverter trip point at the '1' storage node is increased and the SRAM cell can withstand a greater rise at the '0' storage node. Consequently, WRRV has a higher μ , especially at higher supplies, than SRRV. A reduced sensitivity in the pass-gate transistor strength to the WL voltage also implies a larger spread in the WL voltage is needed to produce the same spread in the on-current conducted by the pass-gate transistor. Thus, WRRV measurements display a larger σ than SRRV and exhibit heightened fluctuations in σ compared to SRRV as the supply voltage increases. The shift in the μ/σ value of WRRV compared to SRRV at each supply voltage depends on the relative augmentation in the μ and σ of WRRV. Fig. 13(c) plots the μ/σ value of the measured SRRV and WRRV as a function of the supply voltage. At 0.7 V or below, the μ/σ difference between SRRV and WRRV is small. At 0.8 V and beyond, the μ/σ of WRRV drops below that of SRRV due to a greater augmentation in the σ of WRRV. Measurements of WRRV above 0.9 V were not conducted to keep the WL voltage below 1.5 V to avoid transistor gate-oxide breakdown. Thus, WRRV is not suitable for read stability characterization at higher operating voltages but can be useful for read stability

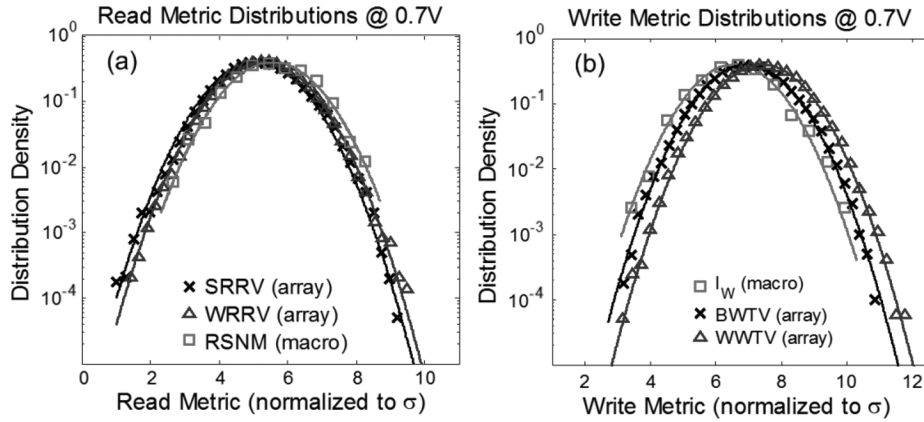


Fig. 12. Measured (a) read metric distributions using RSNM, SRRV, and WRRV; and (b) write metric distributions using I_W , BWTV, and WWTV at 0.7 V.

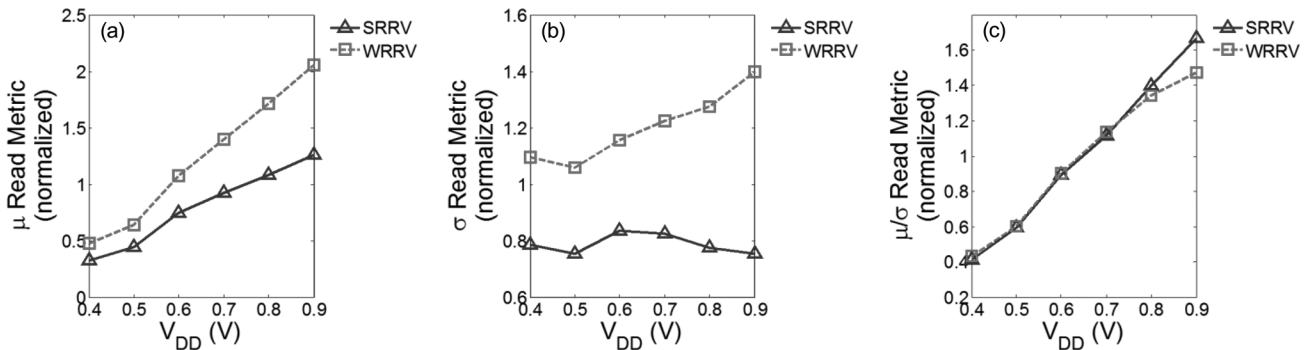


Fig. 13. Measured (a) μ , (b) σ , and (c) μ/σ of SRRV and WRRV as a function of the supply voltage.

characterization near failure. Overall, SRRV can be measured over a more complete range of operating voltages.

Fig. 14(a) and (b) plots the μ and σ of the measured BWTV and WWTV as a function of the supply voltage. Although both BWTV and WWTV measure the write trip voltage of the SRAM cell, their measurements stress the pass-gate transistor differently. During the BL sweep of the BWTV characterization, the strength of the pass-gate transistor at the ‘1’ storage node is modulated through adjusting both its gate-source (V_{GS}) and drain-source voltage (V_{DS}) while the pass-gate transistor at the ‘0’ storage node remains in saturation. Since the source voltage of the pass-gate transistor at the ‘1’ storage node is ramped from V_{DD} to V_{SS} , the pass-gate transistor is first put under reverse body bias (RBB) and the magnitude of the applied RBB drops during the BL sweep. Additionally, the V_{TH} of the pass-gate transistor is modulated by a drain-induced barrier lowering (DIBL) effect during the BL sweep due to a changing V_{DS} . This leads to a varying pass-gate V_{TH} throughout the BWTV characterization process. On the other hand, the WL sweep of the WWTV characterization modulates the strengths of both pass-gate transistors through adjusting only the gate-source (V_{GS}) voltage, leading to a fixed pass-gate V_{TH} until a successful write trip occurs. As a result, the μ of the measured WWTV has a linear dependence on V_{DD} [14], [20] whereas the μ of the measured BWTV is slightly less linearly dependent on V_{DD} [Fig. 14(a)]. In addition, the increased within-die variation of V_{TH} due to RBB [21] and the varying V_{TH} as the degree of

DIBL and RBB varies throughout the BL sweep increases the σ in the BWTV measurements [Fig. 14(b)]. Therefore, WWTV measurements show reduced σ fluctuation as compared with BWTV measurements as the supply voltage changes, due to a fixed pass-gate V_{TH} .⁴ Fig. 14(c) plots μ/σ values of the measured BWTV and WWTV as a function of the supply voltage. The shift in the μ/σ value of BWTV compared to WWTV at each supply voltage value depends on the relative augmentations in the μ and σ of BWTV. At 0.6 V and below, the shift in the μ/σ value is small as BWTV has similar augmentations in μ and σ . At 0.7 V and above, the μ/σ value of BWTV drops below that of WWTV. Up to a 10% shift in the μ/σ value between BWTV and WWTV is observed. Due to a more linear dependence on V_{DD} and a fixed pass-gate V_{TH} , WWTV can more efficiently quantify the impact of pass-gate V_{TH} and supply voltage on SRAM writeability.

2) *Read/Write Metric Correlations:* To investigate the correlations between the various read stability metrics and between the various writeability metrics, scatter plots are generated for WRRV versus RSNM [Fig. 15(a)], SRRV versus WRRV [Fig. 15(b)], WWTV versus I_W [Fig. 16(a)], and BWTV versus WWTV [Fig. 16(b)]. Each pair of metrics in the scatter plots is measured for the same set of SRAM cells first at $V_{DD} = 0.8$ V and then at $V_{DD} = 0.5$ V to expose low read stability and

⁴ σ increases at very low supply voltages (~ 500 mV) for both BWTV and WWTV as transistors enter weak-inversion and subthreshold regions of operation, where $\sigma(V_{TH})$ leads to enhanced variability in the metrics. This also applies for SRRV and WRRV.

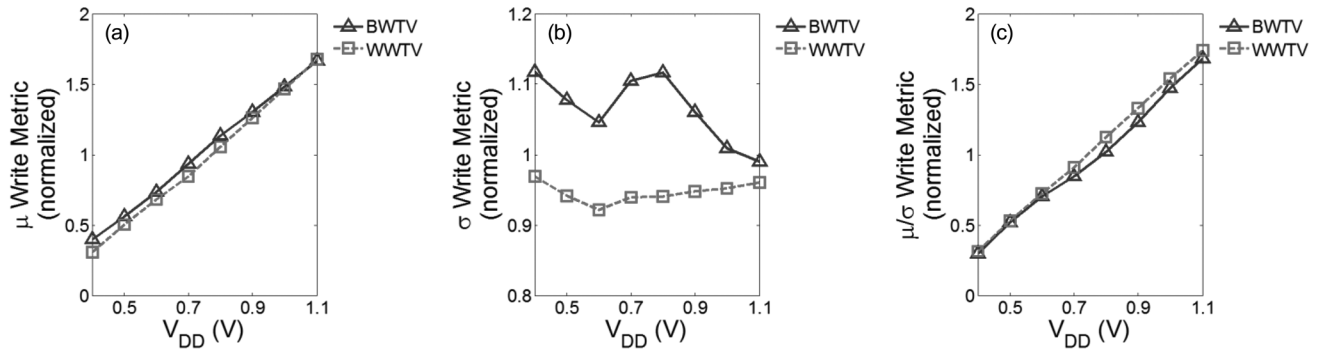


Fig. 14. Measured (a) μ , (b) σ , and (c) μ/σ of BWTV and WWTV as a function of the supply voltage.

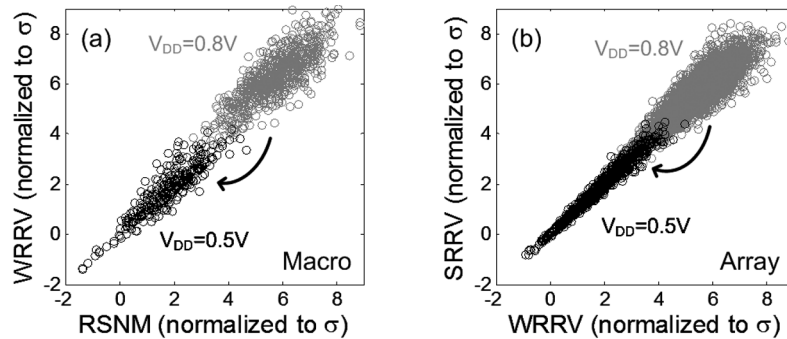


Fig. 15. (a) Scatter plot of WRRV versus RSNM measured from the same SRAM macro with externally accessible storage nodes at 0.8 V and 0.5 V. (b) Scatter plot of SRRV versus WRRV measured from the same functional SRAM array using direct bit line access at 0.8 V and 0.5 V. Excellent correlation is established near the zero crossing.

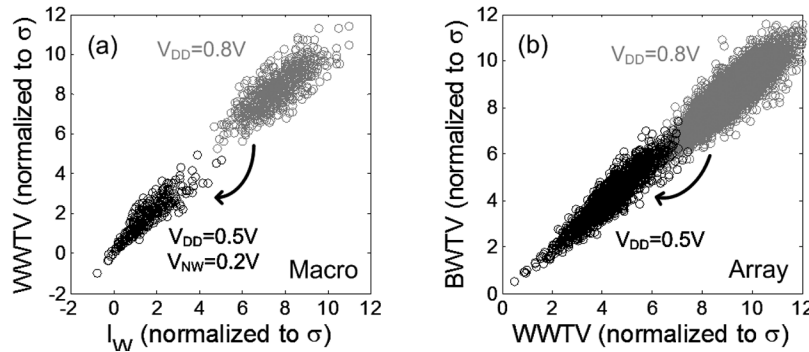


Fig. 16. (a) Scatter plot of WWTV versus I_W measured from the same SRAM macro with externally accessible storage nodes at 0.8 V and at 0.5 V with 200 mV nMOS body bias (V_{BN}). (b) Scatter plot of BWTV versus WWTV measured from the same functional SRAM array using direct bit line access at 0.8 V and 0.5 V. Excellent correlation is established near the zero crossing.

writeability. WRRV-RSNM pairs [Fig. 15(a)] and WWTV- I_W pairs [Fig. 16(a)] are measured from SRAM macros with externally accessible storage nodes; SRRV-WRRV pairs [Fig. 15(b)] and BWTV-WWTV pairs [Fig. 16(b)] are measured from functional SRAM arrays using direct bit-line access. A 200 mV N-well bias (V_{NW}) is applied for the case of $V_{DD} = 0.5$ V when measuring the writeability in the SRAM macros—using WWTV and I_W [Fig. 16(a)]; this is done to further reduce writeability and expose cell failures by decreasing the V_{TH} of the pMOS pull-up transistors. For $V_{DD} = 0.8$ V, the μ of each measured metric sits comfortably above 6σ and a slight dispersion is observed in the measured data of each metric pair. This dispersion is generally smaller at lower measured

values and larger at higher measured values. However, when the supply is dropped⁵ to 0.5 V and the SRAM cell is pushed to the edge of stability, excellent agreement is established within each metric pair, especially near the zero crossing (the origin). This demonstrates that RSNM, SRRV, and WRRV (I_W , BWTV, and WWTV) share the same point of failure—zero crossing—and have excellent agreement near failure, suggesting all metrics can be used for SRAM failure estimation. Furthermore, this also suggests that the μ/σ value, especially taken at higher supply voltages, is highly dependent on the read/write metric used and is therefore unsuitable for estimating yield.

⁵A 200 mV N-well bias (V_{NW}) is also applied for writeability measurements in the SRAM macro.

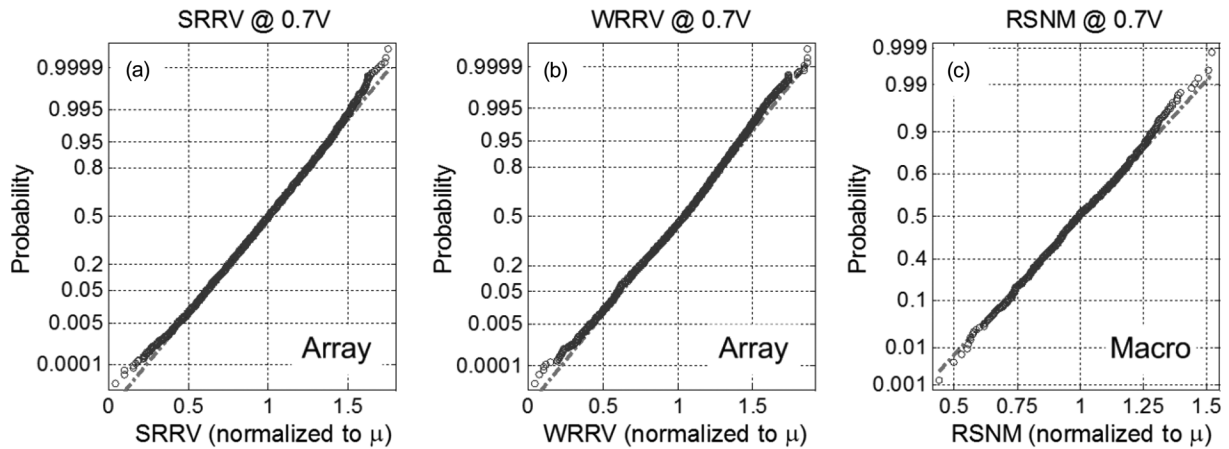


Fig. 17. Normal probability plot for (a) SRRV, (b) WRRV, and (c) RSNM at 0.7 V. SRRV and WRRV are measured from functional SRAM arrays and RSNM is measured from SRAM macros with externally accessible storage nodes.

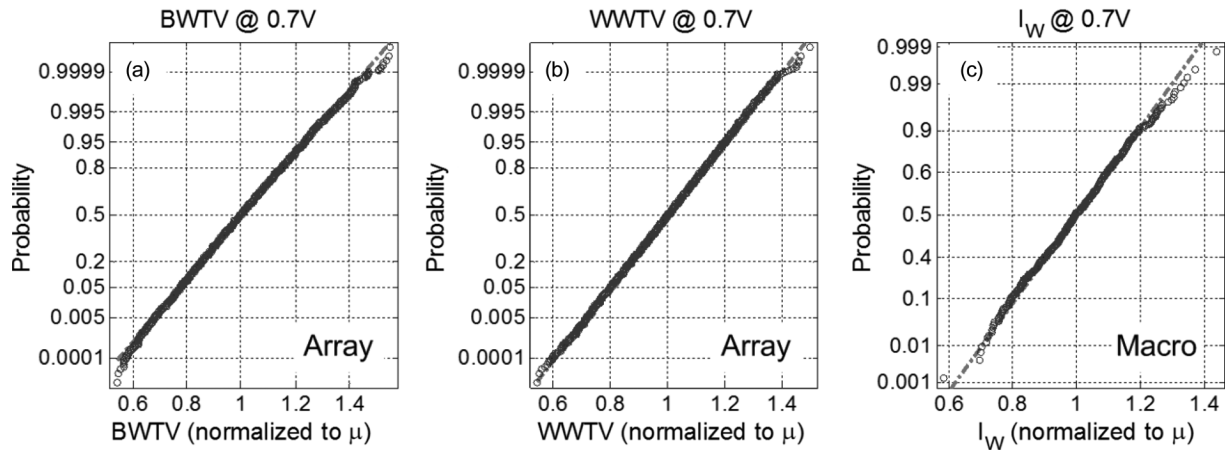


Fig. 18. Normal probability plot for (a) BWTV, (b) WWTV, and (c) I_W at 0.7 V. BWTV and WWTV are measured from functional SRAM arrays and I_W is measured from SRAM macros with externally accessible storage nodes.

3) *Normal Probability Plots*: Fig. 17 shows the normal probability plots [22] for SRRV, WRRV, and RSNM measured at $V_{DD} = 0.7$ V. SRRV and WRRV are measured in 64 kb functional SRAM arrays. These measurements show good normality at the center of the distribution. However, both the upper and the lower tails of the distributions, above and below $\pm 3\sigma$, show slight deviations from a normally distributed function [Fig. 17(a) and (b)]. In both cases, the measured data show a higher probability for a lower read stability value—this is because both SRRV (Fig. 3) and WRRV (Fig. 5) are measured from the less read-stable storage side of the SRAM cell, the equivalent of taking the minimum of two distributions. RSNM is measured in SRAM macros via all-internal-node access characterization for both sides of each SRAM cell⁶ and the measurements show good normality up to $\pm 3\sigma$ [Fig. 17(c)].

Fig. 18 shows the normal probability plot for BWTV, WWTV, and I_W measured at $V_{DD} = 0.7$ V. BWTV and WWTV are measured in 64 kb functional SRAM arrays. These measurements show good normality for writeability values

⁶RSNM in Fig. 17(c) is taken from both sides of each SRAM cell and not the less read-stable side. Therefore, measured RSNM data shows no significant deviation from a normal distribution.

down to more than -4σ [Fig. 18(a) and (b)]. I_W is measured in SRAM macros via all-internal-node access characterization and its measurements show no significant deviation from a normal distribution up to $\pm 3\sigma$ [Fig. 18(c)].

B. Read Current Measurements

Fig. 19(a) and (b) shows the normal probability plot for I_{READ} measured at $V_{DD} = 1.1$ V, $V_{DD} = 0.7$ V, and $V_{DD} = 0.5$ V. At the nominal supply of 1.1 V, I_{READ} has a nearly linear dependence on the V_{TH} of the pass-gate transistor operating in velocity saturation and the pull-down transistor operating in the linear mode; therefore its measurement exhibits good normality up to more than $\pm 4\sigma$ [4]. As the supply is dropped to 0.7 V, some pass-gate transistors are no longer velocity saturated while some pull-down transistors also enter the saturation mode (due to a rise in the ‘0’ storage node and high transistor V_{TH}), causing I_{READ} to no longer linearly depend on the V_{TH} of either pass-gate or pull-down transistors. Thus, the lower tail of the I_{READ} distribution at 0.7 V exhibits a significant right-skewed deviation from a normal distribution [Fig. 19(b)]. When the supply is further dropped to 0.5 V, I_{READ} distribution exhibits even greater right-skewed

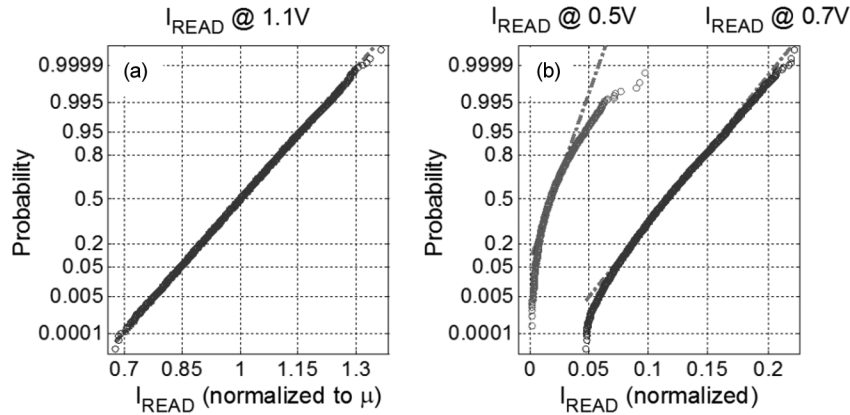


Fig. 19. Normal probability plot for (a) I_{READ} measured at nominal 1.1 V and (b) I_{READ} measured at 0.7 V and 0.5 V from functional SRAM arrays.

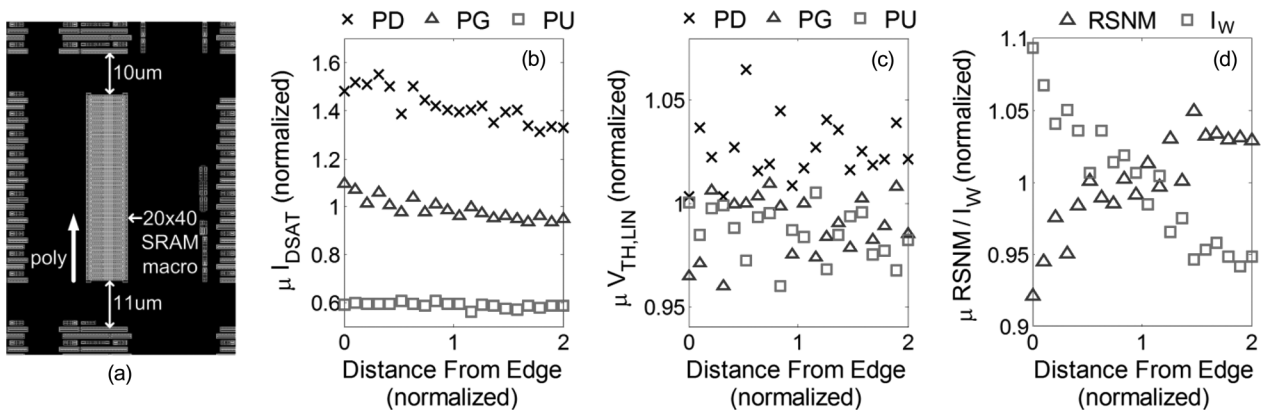


Fig. 20. (a) Layout view of a 20×40 SRAM macro, with poly in the vertical direction, using all-internal-node access surrounded by large STI. (b) Normalized μ of measured I_{DSAT} for pull-down, pass-gate, and pull-up transistors as a function of distance from the edge of the array (normalized to the average distance). (c) Normalized μ of measured $V_{\text{TH,LIN}}$ for pull-down, pass-gate, and pull-up transistors as a function of distance from the edge of the array. (d) Normalized μ of RSNM and I_{W} as a function of distance from the edge of the array. All measurements are taken from SRAM macros via all-internal-node access.

deviations from a normal distribution as transistors enter the brink of weak inversion.

C. Effects of Shallow Trench Isolation (STI) Induced Stress

Fig. 20(a) presents the layout view of a 20-row by 40-column SRAM macro, with poly gates running in the vertical direction, using all-internal-node access and separated from the thick-oxide switch network by wide regions of shallow trench isolation (STI) on all four sides. Fig. 20(b) and (c) summarize the impact of STI induced stress on transistor performance in this low-power 45 nm process. Transistor channels in this process are oriented in the $\langle 100 \rangle$ direction, making the pMOS transistors insensitive to stress while enhancing hole mobility. This process uses sub-atmospheric chemical vapor deposition (SACVD) for trench isolation. As a result, the trenches exert a weak tensile strain orthogonal (transverse) [23] to the direction of current flow on the nMOS transistors rather than a strong compressive strain [16], [18]. Fig. 20(b) reveals a systematic decrease in the I_{DSAT} of the nMOS pass-gate and pull-down transistors away from the periphery of the SRAM macro while the I_{DSAT} of the pMOS pull-up transistor is unaffected. The recorded drop in I_{DSAT} from the edge to the center of the macro

was roughly 10%. In addition, $V_{\text{TH,LIN}}$ of both nMOS and pMOS transistors display no dependence on their distance from the edge of the SRAM macro [Fig. 20(c)]. This confirms an enhancement in the nMOS transistor mobility due to the tensile strain induced by the large STI regions peripheral to the SRAM macro. Fig. 20(d) illustrates the impact of STI induced stress on SRAM read stability and writeability. Due to a decrease in the nMOS transistor strength while the pMOS transistor strength stays unaffected, RSNM of the SRAM cells increases away from the STI interface of the SRAM macro while the I_{W} drops. Due to a more direct impact of nMOS to pMOS transistor ratio on the writeability of the cell, the drop in I_{W} from edge to center of the macro was roughly 15% whereas the rise in RSNM was just over 10%.

Fig. 21 plots the measured SRRV, WWTV,⁷ and I_{READ} as a function of row and column position within a 256×256 functional SRAM array. Although significant systematic shifts in the SRAM RSNM and I_{W} were observed in the SRAM macros with all-internal-node access, no such systematic drifts were observed in SRRV, WWTV, and I_{READ} measured in functional SRAM arrays as they are densely surrounded by peripheral circuitry. All measured metrics vary randomly with position within

⁷WWTV in Fig. 21(b) is taken from the CL side of each SRAM cell.

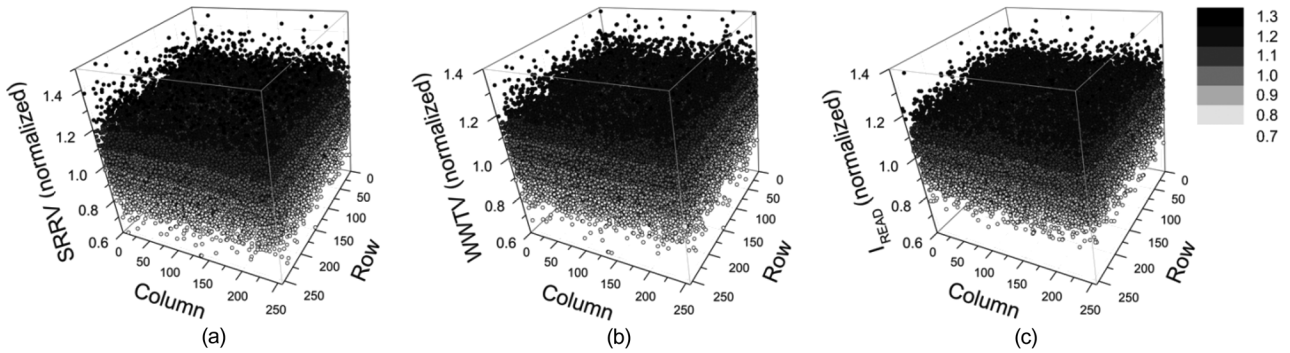


Fig. 21. Measured (a) SRRV, (b) WWTV, and (c) I_{READ} as a function of row and column position within a 256×256 functional SRAM array.

each 64 kb SRAM array (Fig. 21). This emphasizes the importance of characterizing the stability of SRAM cells in their natural environment.

D. Within-Cell Mismatch and Cell Orientation

SRAM cells are typically mirrored both horizontally and vertically to maximize the array density, yielding 4 different cell orientations [Fig. 22(a)]. In this 4-cell cluster, orientations A and D share the same layout, with reversed storage nodes; likewise, orientations B and C share the same layout, with reversed storage nodes.⁸ Fig. 22(c)–(e) summarizes the effect of within-cell mismatch and cell orientation on SRAM stability and performance. The measurement results from two test chips scattered across the same wafer are highlighted. Since within-cell mismatch causes (at least) one side of the SRAM cell to flip when V_{DD} is dropped or WL is overdriven, the frequency of read disturbance as a function of cell storage node and cell orientation is plotted in Fig. 22(c). Fig. 22(d) and (e) plots the normalized μ of the measured WWTV and I_{READ} as a function of cell storage node and orientation. The test chip locations within the wafer are identified in Fig. 22(b). Measurement data reveals up to 4X difference in read disturb frequency, 4% shift in the μ of measured WWTV, and 8% shift in the μ of measured I_{READ} when the data polarity of the cell is changed. The shift is consistent throughout the chip, suggesting a systematic mismatch between the two halves of the SRAM cell which may be attributed to a difference in the direction of the notches in the nMOS active region [24]–[26]. The direction of this systematic mismatch can be the same or opposite between two test chips taken from the same wafer, as highlighted in Fig. 22(c)–(e). The measurement data also indicates that the directions of the shifts in read disturb frequency, WWTV, and I_{READ} are correlated—i.e., a higher read disturb frequency typically corresponds to a higher writeability and I_{READ} . As the SRAM cells are mirrored across the direction of the poly (from A/D to B/C), a slight alteration in the degree of the within-cell mismatch is observed, suggestive of a slight poly gate to active source/drain misalignment [24], [25]. Since the nMOS active region is continuous throughout the array whereas the pMOS active region is segmented, the channel width of the pMOS transistors is expected to fluctuate more with

⁸Storage nodes are labeled such that the CL and CH side of orientation A (B) have identical layouts as the CL and CH side of orientation D (C).

poly gate misalignment than nMOS transistors, due to corner rounding effects. As the poly lines are shifted to the right or left, both nMOS pull-down and pass-gate transistors on the same storage side of the SRAM cell experience either a common increase or a common decrease in channel width, due to rounding at the corners of the active region, depending on the cell orientation. As long as the degree of poly gate misalignment stays fairly uniform throughout the SRAM array, the fluctuations in the cell β -ratio (i.e., the strength ratio of pull-down to pass-gate transistors) should be small. Therefore, the observed alteration in the degree of within-cell mismatch (between orientations A/D and orientations B/C) is the smallest for the read disturb frequency. This alteration is slightly greater for I_{READ} , which has a more direct dependence on the pass-gate drive strength than the pull-down drive strength; and it is most pronounced in the cell writeability, which directly depends on the strength of the pMOS pull-up transistor.

E. V_{MIN} Measurements

1) *Distributions and Correlations With Read/Write Metrics:* Fig. 23 plots the V_{MIN} distributions for static read and write operations measured in a 64 kb SRAM array. In both cases, V_{MIN} measurements are extracted from the less stable side of each SRAM cell—i.e., the maximum value of V_{MIN} is taken for each cell. Fig. 23(a) shows that many SRAM cells in the array can achieve read data retention at very low operating voltages—well into the subthreshold region of operation, and thus yield a log-normal shaped distribution due to the exponential dependence of transistor currents on V_{TH} in this region. However, the first read cell bit failure occurs at just over $0.6 \times V_{\text{DD,NOMINAL}}$. Similarly, Fig. 23(b) shows that the first write cell bit failure also occurs at slightly over $0.6 \times V_{\text{DD,NOMINAL}}$. Weak write [27], in the form of a 100 mV reduction in the WL voltage,⁹ is applied to expose a reasonable amount of failure above subthreshold due to the high writeability of SRAM cells tested. The upper tail of the write V_{MIN} distribution fits well to a Gaussian distribution as SRAM cells still operate above threshold. As SRAM cells enter the deep subthreshold region of operation, the pass-gate transistors cannot access the SRAM cells and the bit flip is caused by a standby retention failure—yielding a high cell bit failure count

⁹A reduced WL voltage is applied rather than a raised BL voltage (at the ‘1’ storage node) to preserve a constant pass-gate V_{TH} .

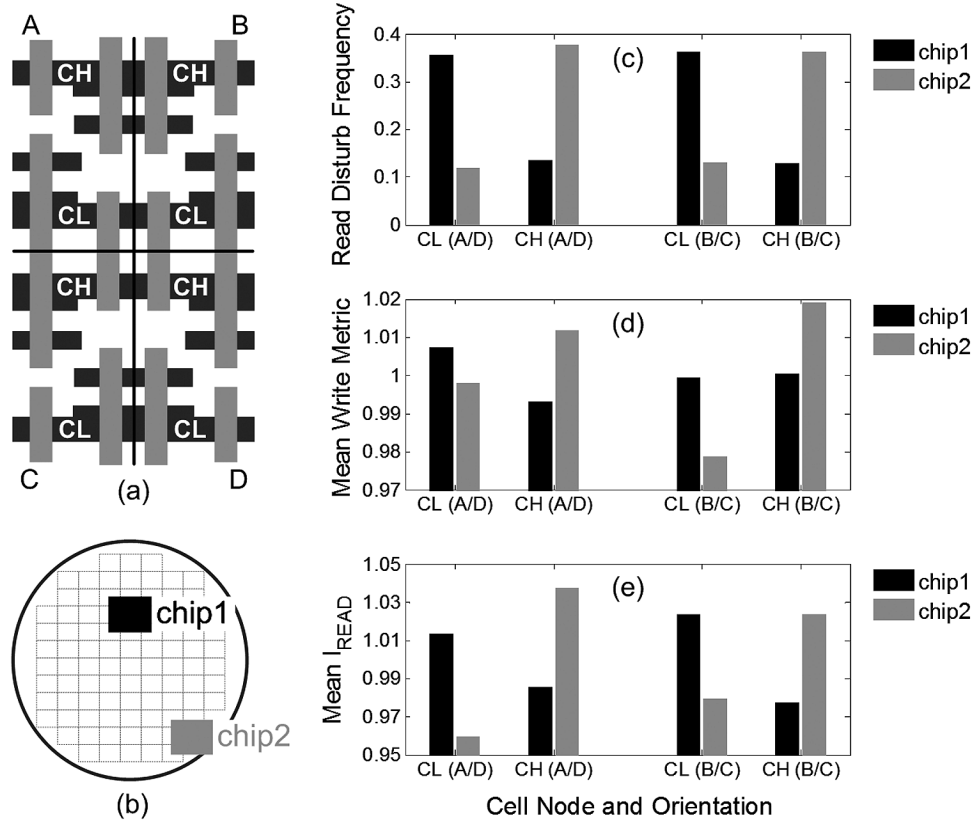


Fig. 22. (a) Four-cell cluster in an SRAM array showing 4 cell orientations; the storage nodes of orientations C and D are reversed in the drawing for clarification. (b) Wafer map identifying the measured chips. Measured (c) read disturb frequency, (d) normalized μ of WWTV, and (e) normalized μ of I_{READ} for two test chips on the same wafer as a function of cell storage node and orientation.

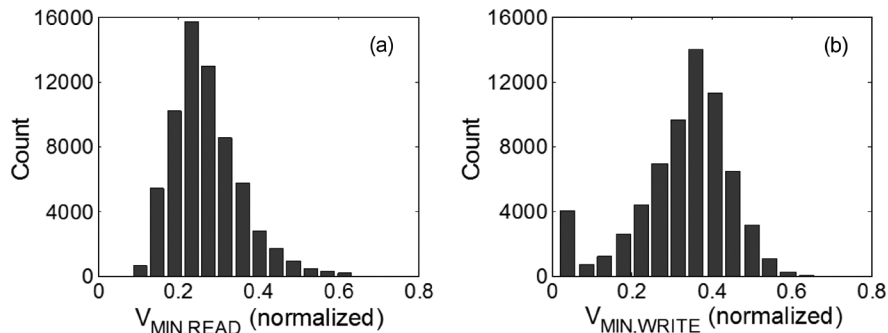


Fig. 23. Distributions of (a) V_{MIN} during a static read operation and (b) V_{MIN} during a static write operation measured in a 64 kb SRAM array. All V_{MIN} values are normalized to the nominal supply voltage ($V_{\text{DD,NOMINAL}}$).

near 0 V.¹⁰ Since SRAM yield is limited by the first few cell bit failure occurrences, only the upper ends of both distributions are significant.

A direct correlation between extracted V_{MIN} and SRAM read/write margin measurements from a functional SRAM array is plotted in Fig. 24. Fig. 24(a) shows the scatter plot of SRRV versus $V_{\text{MIN,READ}}$, where SRRV is measured at 0.6 V to expose near-failure read stability. The scatter plot of WWTV (measured at 0.6 V) versus $V_{\text{MIN,WRITE}}$ (measured with a 100 mV WL weak write) is shown in

¹⁰Retention fails cannot be easily distinguished from write fails under very low operating voltages. This is generally not an issue because only the upper tail of the write V_{MIN} distribution is important.

Fig. 24(b). Results show excellent agreement between $V_{\text{MIN,READ}}/V_{\text{MIN,WRITE}}$ and SRRV/WWTV measurements, especially at high $V_{\text{MIN,READ}}/V_{\text{MIN,WRITE}}$ and low SRRV/WWTV values. Since SRRV is measured at 0.6 V, the zero crossing of the SRRV measurements, in Fig. 24(a), corresponds to $V_{\text{MIN,READ}} = 0.6$ V (before normalization).¹¹ Fig. 24(b) shows a large cloud near the y axis at high WWTV and low V_{MIN} values, corresponding to the V_{MIN} values captured due to standby retention failures. The excellent agreement between extracted V_{MIN} and large-scale read/write margin measurements suggest that read/write metrics measured using

¹¹The zero crossing of the WWTV measurements does not correspond to $V_{\text{MIN,WRITE}} = 0.6$ V due to the application of a 100 mV WL weak write.

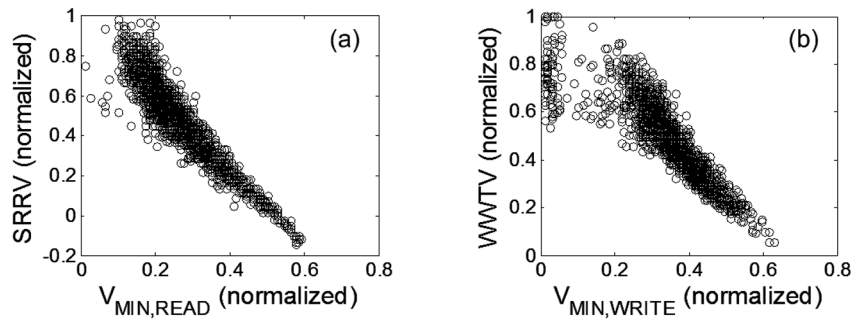


Fig. 24. Scatter plots of (a) SRRV versus V_{MIN} during a static read operation and (b) WWTV versus V_{MIN} during a static write operation (using a 100 mV WL weak write) measured in a 64 kb SRAM array. All V_{MIN} values are normalized to $V_{DD,NOMINAL}$. Both SRRV and WRRV are measured at 0.6 V to establish correlation near the zero crossing.

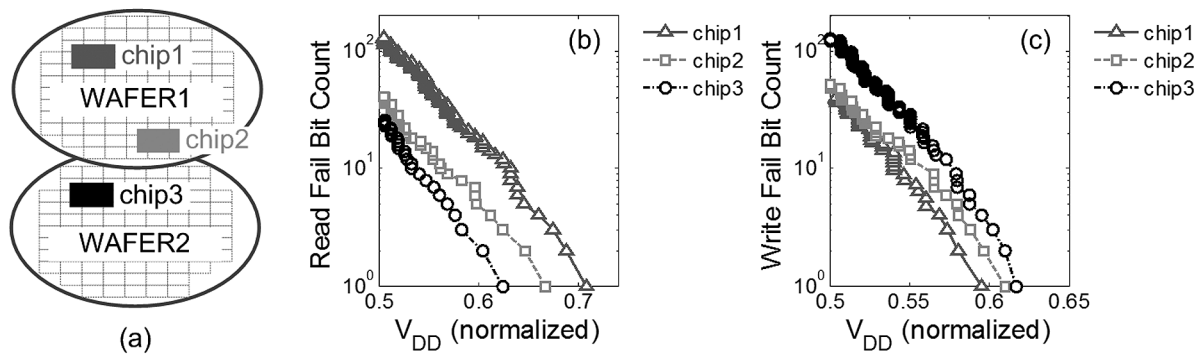


Fig. 25. (a) Locations of measured test chips on two different wafers. (b) Fail bit count as a function of normalized V_{DD} during a static read operation and (c) fail bit count as a function of normalized V_{DD} during a static write operation measured for SRAM arrays on three test chips from two different wafers.

direct bit-line characterization, which are easier to model than V_{MIN} and can be fitted to simple Gaussian distributions, can be used to estimate V_{MIN} .

2) *Within-Wafer and Wafer-to-Wafer V_{MIN} Variations:* Fig. 25 plots the fail bit count as a function of the normalized V_{DD} for static read and write operations measured for a 64 kb SRAM array from three test chips separately located on two different wafers. The locations of the three test chips are identified in Fig. 25(a)—chip1 and chip2 are scattered across wafer1 and chip3 shares the same location as chip1 on wafer2. The two wafers have a nominal 4 nm difference in the effective transistor channel length (L_{EFF}), corresponding to different process corners [16], where wafer1 represents the faster wafer. A 100 mV WL weak write is applied during the $V_{MIN,WRITE}$ measurements for all three test chips. Results reveal a within-wafer systematic shift in the measured V_{MIN} as chip1 shows a 3% reduction in the write-fail free V_{DD} and a 6% climb in the read-upset free V_{DD} compared to chip2. A notable wafer-to-wafer systematic shift in the measured V_{MIN} is also observed as chip3 displays a 5% increase in the write-fail free V_{DD} and a 9% drop in the read-upset free V_{DD} compared to chip1. In addition to a shift in the L_{EFF} , measurements of transistor currents from the SRAM macros with all-internal-node access confirmed that the pMOS transistors on wafer1 have degraded current drive, due to a higher V_{TH} , compared to pMOS transistors on wafer2, leading to enhanced writeability and degraded read stability.

3) *Impact of Read Assist Circuits (RAC) and Write Assist Circuits (WAC):* Four conventionally used read assist circuit (RAC) and write assist circuit (WAC) schemes are applied to a 64 kb SRAM array to improve the read stability and writeability. The corresponding improvements in $V_{MIN,READ}$ and $V_{MIN,WRITE}$ are compared in Figs. 26 and 27. Fig. 26(a) shows the simple schematics of the boosted cell V_{DD} (BCV) scheme [24], [28], [29] for read stability enhancement and the cell V_{DD} down (CVD) scheme [28] for writeability enhancement. These schemes can be combined to simultaneously enhance read stability and writeability—by boosting the cell supply during the read cycle and suppressing the cell supply during the write cycle—for SRAM cells utilizing the thin-cell topology with column-based supply routing [28]. Fig. 26(b) and (c) plots the fail bit count as a function of the normalized V_{DD} for static read and write operations before and after applying a 100 mV BCV and a 100 mV CVD. Alternatively, SRAM read stability can be enhanced through the suppressed word-line (SWL) scheme [25], [30] and the negative bit-line (NBL) scheme [30]–[32] can be applied to enhance the SRAM writeability [Fig. 27(a)]. Fig. 27(b) and (c) plots the fail bit count as a function of the normalized V_{DD} for static read and write operations before and after applying a 100 mV SWL and a 100 mV NBL. Measurements reveal that a 100 mV BCV and a 100 mV SWL achieve similar $V_{MIN,READ}$ enhancements—25% and 24% respectively; however, applying a 100 mV SWL may negatively impact the read access per-

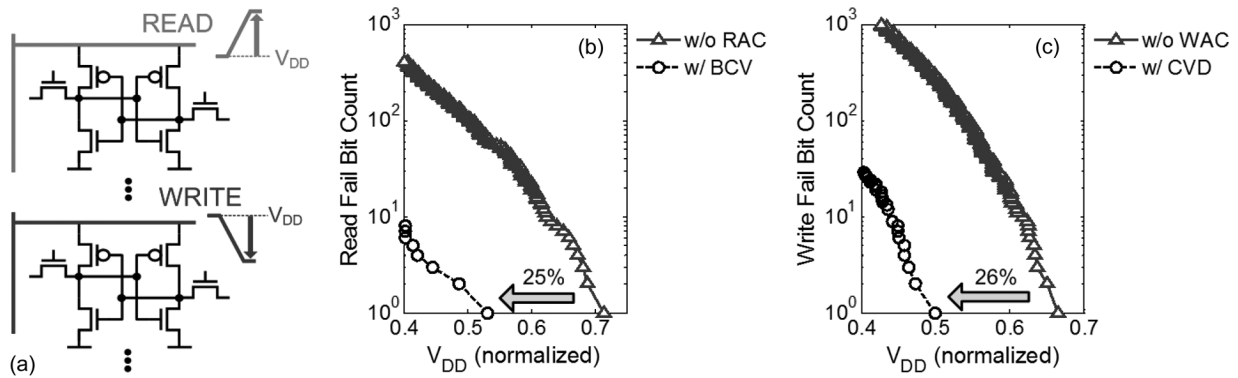


Fig. 26. (a) Simplified schematic of the boosted cell V_{DD} (BCV) and the cell V_{DD} down (CVD) scheme for read and write assist. (b) Fail bit count as a function of normalized V_{DD} during a read operation measured for the same SRAM array with no read assist circuits (RAC) and with a 100 mV BCV. (c) Fail bit count as a function of normalized V_{DD} during a write operation measured for the same SRAM array with no write assist circuits (WAC) and with a 100 mV CVD.

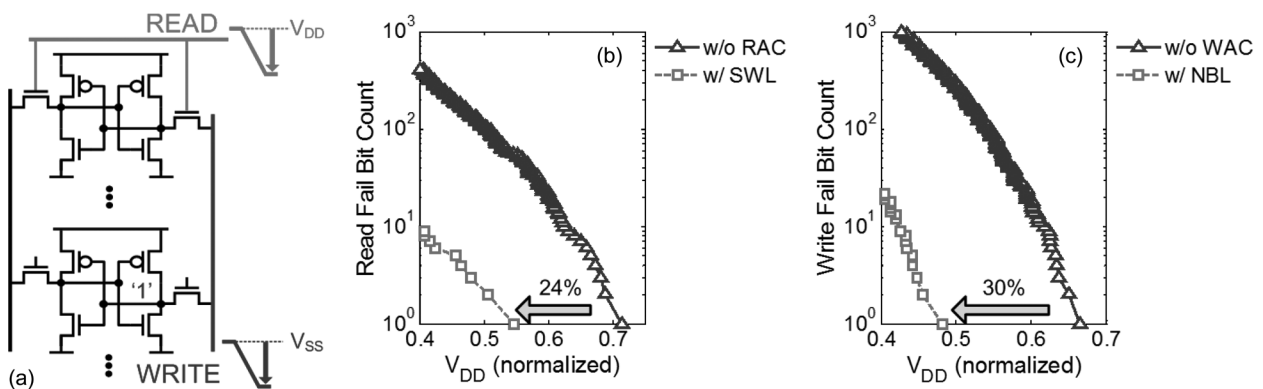


Fig. 27. (a) Simplified schematic of the suppressed word-line (SWL) and the negative bit-line (NBL) scheme for read and write assist. (b) Fail bit count as a function of normalized V_{DD} during a read operation measured for the same SRAM array with no read assist circuits (RAC) and with a 100 mV SWL. (c) Fail bit count as a function of normalized V_{DD} during a write operation measured for the same SRAM array with no write assist circuits (WAC) and with a 100 mV NBL.

formance due to a degraded gate overdrive of the pass-gate transistor during the read cycle. On the other hand, a 100 mV NBL achieves slightly better $V_{MIN,WRITE}$ enhancements than a 100 mV CVD—30% and 26% respectively—because applying an nMOS pull-down gate-source overdrive can more effectively adjust the cell α ratio (i.e., pMOS to nMOS ratio) than suppressing the pMOS pull-up drain-source bias [30]. Additionally, the NBL technique remains effective for writeability enhancement for dual-port SRAM, whereas CVD degrades the read stability of a simultaneously accessed cell from the same column [30], [32]. However, a small positive V_{GS} is applied for all unaccessed SRAM cells in the same column when NBL is activated; this may lead to increased BL leakage current during write cycles. The scalability of the NBL technique may also be limited, due to a reduction in the maximum tolerable voltage across the gate oxide as its thickness scales down.¹²

VII. CONCLUSION

A direct bit-line measurement methodology to characterize SRAM cell read stability and writeability in functional SRAM arrays is demonstrated at the 45 nm technology node. This methodology is further extended for the characterization of

¹²The same limitation applies to BCV, since boosting the cell V_{DD} increases the voltage across the gate oxide of the pMOS pull-up transistor.

SRAM V_{MIN} during read and write cycles. The large-scale read/write margin measurements show excellent correlation, near failure, to SRAM DC RSNM and I_W measurements conducted in small SRAM macros with all-internal-node access. In addition, a direct correlation between measured SRAM read/write margins and V_{MIN} in a functional SRAM array is established. Results demonstrate excellent agreement between large-scale read/write metrics and V_{MIN} , especially at low read/write margin and high V_{MIN} regions; this suggests that large-scale read/write metrics can be used for V_{MIN} estimation. Several sources of process-induced systematic mismatch are identified using large-scale measurements and investigated. The impacts of four conventionally used read assist and write assist techniques on SRAM V_{MIN} are evaluated and compared. The characterization of SRAM stability in large functional arrays is capable of delivering substantial statistics for SRAM failure analysis. Moreover, this methodology can be easily extended to capture more than 6 standard deviations of parameter variations by increasing the SRAM array size.

REFERENCES

- [1] A. Bhavnagarwala *et al.*, "Fluctuation limits and scaling opportunities for CMOS SRAM cells," in *IEDM Tech. Dig.*, 2005, pp. 675–678.
- [2] A. Bhavnagarwala *et al.*, "A sub-600 mV, fluctuation tolerant 65 nm CMOS SRAM array with dynamic cell biasing," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 946–955, Apr. 2008.

- [3] Z. Guo, A. Carlson, L.-T. Pang, K. Duong, T.-J. King, and B. Nikolić, "Large-scale read/write margin measurement in 45 nm CMOS SRAM arrays," in *Symp. VLSI Circuits Dig.*, 2008, pp. 42–43.
- [4] T. Fischer *et al.*, "Analysis of read current and write trip voltage variability from a 1-MB SRAM test structure," *IEEE Trans. Semicond. Manufact.*, vol. 21, no. 4, pp. 534–541, Nov. 2008.
- [5] M. Agostinelli *et al.*, "Erratic fluctuations of SRAM cache V_{MIN} at the 90 nm process technology node," in *IEDM Tech. Dig.*, 2005, pp. 655–658.
- [6] M. Ball *et al.*, "A screening methodology for V_{MIN} drift in SRAM arrays with applications to sub-65 nm nodes," in *IEDM Tech. Dig.*, 2006, pp. 1–4.
- [7] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987.
- [8] C. Wann *et al.*, "SRAM cell design for stability methodology," in *VLSI-TSA Dig.*, 2005, pp. 21–22.
- [9] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2577–2588, Nov. 2006.
- [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] Rashmi, A. Kranti, and G. A. Armstrong, "6-T SRAM cell design with nanoscale double-gate SOI MOSFETs: Impact of source/drain engineering and circuit topology," *IOP Semiconductor Science and Technology*, vol. 23, Jul. 2008.
- [12] V. K. Wong, C. H. Lock, K. H. Siek, and P. J. Tan, "Electrical analysis to fault isolate defects in 6T memory cells," in *Proc. IEEE IPFA*, 2002, pp. 101–104.
- [13] R. Heald and P. Wang, "Variability in sub-100 nm SRAM designs," in *Proc. IEEE/ACM ICCAD*, 2004, pp. 347–352.
- [14] N. Gierczynski, B. Borot, N. Planes, and H. Brut, "A new combined methodology for write-margin extraction of advanced SRAM," in *Proc. IEEE ICMITS*, 2007, pp. 97–100.
- [15] A. Carlson, Z. Guo, L.-T. Pang, T.-J. King, and B. Nikolić, "Compensation of systematic variations through optimal biasing of SRAM word-lines," in *Proc. IEEE CICC*, 2008, pp. 411–414.
- [16] L.-T. Pang, K. Qian, C. Spanos, and B. Nikolić, "Measurement and analysis of variability in 45 nm strained-Si CMOS technology," *IEEE J. Solid-State Circuits*, vol. 44, no. 8, pp. 2233–2243, Aug. 2009.
- [17] E. Josse *et al.*, "A cost-effective low power platform for the 45-nm technology node," in *IEDM Tech. Dig.*, 2006, pp. 1–4.
- [18] C. Le Cam *et al.*, "A low cost drive current enhancement technique using shallow trench isolation induced stress for 45-nm node," in *VLSI Tech. Dig.*, 2006, pp. 82–83.
- [19] B. Le Gratiet *et al.*, "Process control for 45 nm CMOS logic gate patterning," *Proc. SPIE*, vol. 6922, Mar. 2008.
- [20] K. Takeda, H. Ikeda, Y. Hagihara, M. Nomura, and H. Kobatake, "Re-definition of write margin for next-generation SRAM and write-margin monitoring circuit," in *IEEE ISSCC Dig.*, 2006, pp. 2602–2611.
- [21] S. Narendra, D. Antoniadis, and V. De, "Impact of using adaptive body bias to compensate die-to-die V_t variation on within-die V_t variation," in *Proc. IEEE ISLPED*, 1999, pp. 229–232.
- [22] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth, 1983.
- [23] S. E. Thompson *et al.*, "A logic nanotechnology featuring strained-silicon," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 191–193, Apr. 2004.
- [24] M. Yamaoka, K. Osada, and K. Ishibashi, "0.4 V logic-library-friendly SRAM array using rectangular-diffusion cell and delta-boosted-array voltage scheme," *IEEE J. Solid-State Circuits*, vol. 39, no. 6, pp. 934–940, Jun. 2004.
- [25] S. Ohbayashi *et al.*, "A 65 nm SoC embedded 6T-SRAM design for manufacturing with read and write cell stabilizing circuits," in *VLSI Circuits Dig.*, 2006, pp. 17–18.
- [26] M. Khellah *et al.*, "PVT-variations and supply-noise tolerant 45 nm dense cache arrays with Diffusion-Notch-Free (DNF) 6T SRAM cells and dynamic multi-Vcc circuits," in *VLSI Circuits Dig.*, 2008, pp. 48–49.
- [27] A. Meixner and J. Banik, "Weak write test mode: An SRAM cell stability design for test technique," in *Proc. ITC*, 1997, pp. 1043–1052.
- [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] Y. Chung and S.-W. S, "An experimental 0.9 V 256-kbit SRAM macro with boosted cell array scheme," *ETRI J.*, vol. 29, no. 4, pp. 457–462, Aug. 2007.
- [30] K. Nii *et al.*, "A 45-nm single-port and dual-port SRAM family with robust read/write stabilizing circuitry under DVFS environment," in *VLSI Circuits Dig.*, 2008, pp. 212–213.
- [31] 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.
- [32] D. P. Wang *et al.*, "A 45 nm dual-port SRAM with write and read capability enhancement at low voltage," in *Proc. IEEE SOC*, 2007, pp. 211–214.



Zheng Guo (S'03) received the B.S. degree in computer engineering from the University of Illinois at Urbana-Champaign in 2003. Upon graduation, 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. In 2005, he received the M.S. degree in electrical engineering from the University of California at Berkeley, where he continues to work towards the Ph.D. degree. His research emphasis is on variability characterization and robust design techniques for nanoscale SRAM.

In 2004, Mr. Guo 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.

Andrew Carlson (S'04–M'08) received the S.B. degree in engineering sciences from Harvard University in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 2005 and 2008, respectively.

He is currently working at Advanced Micro Devices on technology development for the 22-nm node.



Liang-Teck Pang (S'02–M'08) received the Dipl.Ing. from Ecole Centrale de Paris in France and the Master of Philosophy (M.Phil.) from Cambridge University, U.K, in 1997. Between 1998 and 2002, he worked in the DSO National Labs in Singapore on VLSI implementation of signal processing algorithms and high performance micro-architecture and circuit design. In 2002, he started the Ph.D. program in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. In the summers of 2005–2007, he interned

in IBM Austin Research Labs where he was an inventor on several US patent applications. He was presented with the IBM Invention Achievement Award in recognition of those contributions. Upon completion of his Ph.D. program in 2008, he joined IBM T. J. Watson Research Center as a research staff member.

His Ph.D. research involved the design of circuits to measure and characterize CMOS performance variability due to fluctuations in the manufacturing process. His emphasis is on the effects of layout on CMOS performance, measurement of the spatial correlation of logic gates, and analysis of variability data.

Kenneth T. Duong received the B.S. and M.S. degrees in electrical engineering from the University of California at Berkeley in 2007 and 2008, respectively.

He is currently working at Sun Microsystems on design technology development for the 28-nm node.



Tsu-Jae King Liu (SM'00–F'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1984, 1986, and 1994, respectively.

In 1992, she joined the Xerox Palo Alto Research Center, Palo Alto, CA, as a Member of Research Staff, to research and develop polycrystalline-silicon thin-film transistor technologies for high-performance flat-panel displays. In August 1996, she joined the faculty of the University of California at Berkeley, where she is currently the Conexant

Systems Distinguished Professor of Electrical Engineering and Computer Sciences and Associate Dean for Research in the College of Engineering. Her research activities are presently in nanometer-scale integrated-circuit devices and technology, as well as materials, processes, and devices for integrated microsystems.

Dr. Liu's awards include the DARPA Significant Technical Achievement Award (2000) for development of the FinFET, and the IEEE Kiyu Tomiyasu Award (2010) for contributions to nanoscale MOS transistors, memory devices, and MEMs devices. She has served on committees for many technical conferences, including the International Electron Devices Meeting and the Symposium on VLSI Technology. She was an Editor for the IEEE ELECTRON DEVICES 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 coauthor of the book *Digital Integrated Circuits: A Design Perspective* (2nd ed, 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 Award at the ACM/IEEE International Symposium of Low-Power Electronics in 2005, and the 2004 Jack Kilby Award for the Outstanding Student Paper at the IEEE International Solid-State Circuits Conference.