Programmable Techniques for Cell Stability Test and Debug in Embedded SRAMs

Andrei Pavlov[1], Manoj Sachdev[1], José Pineda de Gyvez[2] and Mohamed Azimane[2]
[1]Department of Electrical & Computer Engineering, University of Waterloo, ON, Canada
[2]Philips Research Labs, Eindhoven, The Netherlands

***Abstract*— Reliable cell stability test of modern embedded SRAMs calls for DFT techniques with a flexible detection threshold. We present two programmable cell stability test and debug techniques that use partially discharged floating bit lines to apply a weak overwrite stress to a cell under test. The applied stress can be digitally adjusted to track the process variations or the desired pass/fail threshold. The proposed techniques are demonstrated to exceed the regular Data Retention Test in both the defect coverage and detection range.**

I. Introduction and Motivation

As technology continues to scale, the increasing process spreads, intra-die threshold voltage mismatches and the rising subtle defect density may result in inadequately low Static Noise Margin (SNM) [1],[2] and cause Stability Faults (SF) in an SRAM cell (Fig. 1(a)). Embedded memory is dominated by SRAM and can occupy the majority of SoC area. Reducing stability of embedded SRAMs becomes a yield limiter in SOCs [3]. As a guideline, $\mu - 6\sigma$ of SNM is required to exceed $\simeq$4% of V_{DD} to reach a 90% yield on 1Mb SRAM [4]. Typically, that translates into a requirement that $SNM_{min} \geq 20\% SNM_{typ}$. Cells with marginally smaller SNM_{min} may escape the traditional tests and fail in the field. Depending on the degree of the SNM degradation, cell stability problems can be classified into Data Retention Faults (DRF) and SF (Fig. 1(c)top). DRFs are the most severe SFs and often caused by missing or poorly formed PMOS transistors (Q3, Q4) in the cell and can be represented by R1 (symmetric defect) and R2, R3 (asymmetric defects) (Fig. 1(b)). If $I_{off\_Q2} > I_{pull-up} + I_{off\_Q6}$, then after a delay proportional to $C_B V_B / (I_{off\_Q2} - (I_{pull-up} + I_{off\_Q6}))$ the capacitance of the node B (C_B) will discharge sufficiently for the cell to flip states. Reading the cell data after a delay on the order of 100ms and comparing it with the previously written data can detect resistive defects R1-R3 in the range of several GΩ. This algorithm is employed by the traditional passive Data Retention Test (DRT, a.k.a. Delay or Pause Test). Positive temperature dependence of the leakage current reduces the SNM and helps to expand the detection range of the DRT to tens of MΩ at 150^0C (Fig. 1(d)). However, opens in this resistance range are still considered as *hard opens* caused by serious defects in the pull-up path of the cell. Open defects with marginally smaller resistance will not be detected by the DRT and such SoCs with defective and potentially unstable and unreliable *weak cells* SRAM cells will be shipped to the customer. Moreover, significant test time and elevated temperatures negatively impact the tester throughput affecting the economics of the DRT.

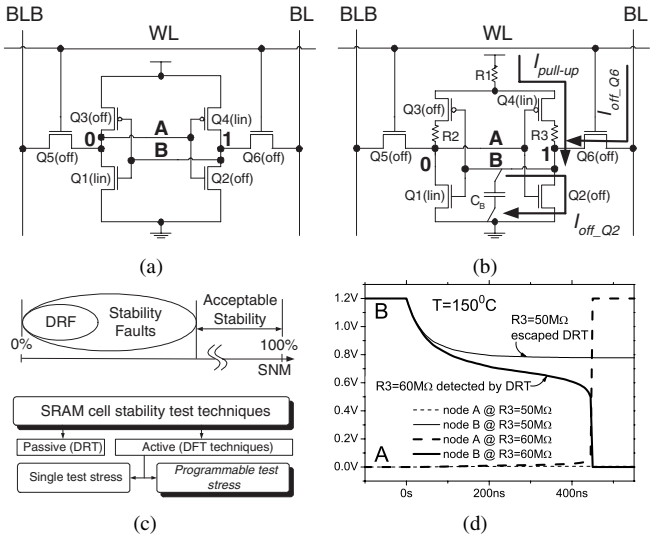


Fig. 1. (a) Defect-free SRAM cell and (b) SRAM cell with a symmetric (R1) and asymmetric (R2, R3) defects in data retention mode; (c) Relationship between the SNM, data retention and stability faults in SRAM cell (top) and classification of the stability test techniques (bottom); (d) Resistive open <50MΩ is not detected by the DRT even at T=150^{o}C.

Reliable and economical detection of weak SRAM cells requires dedicated test circuitry such as Weak Write Test Mode (WWTM) [5] or Soft Defect Detection (SDD) [6]. Beside the defects causing DRFs, such techniques can detect any other defect or excessive process shift that cause degradation of the cell stability (i.e. SNM) below its pass/fail threshold. On one hand, the applied test stress should be strong enough to flip a defective cell and on the other hand - should not flip a healthy cell. Due to poor tracking of process changes/modifications, DFTs with a single test stress (see Fig. 1(c)), such as the WWTM, can over- or under-test the Cell Under Test (CUT) and may require several post-silicon design iterations to adjust the test stress (i.e. pass/fail threshold) accordingly. Whereas, DFTs with programmable test stress allow to adjust the pass/fail threshold during the test without having to modify the DFT circuit. Recently, several programmable techniques for cell stability test have been disclosed. Short Write Test Mode [7] applies a programmable delay element to generate a write pulse short enough not to flip a healthy cell. The Programmable WWTM disclosed in [8] uses dedicated decoder and bias generator to supply

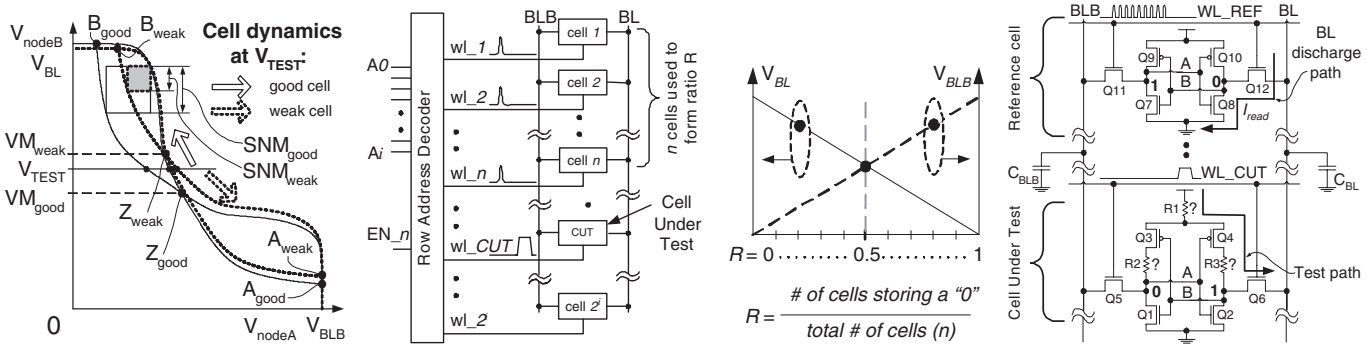


Fig. 2. (a) Choice of V_{TEST} with respect to the metastable points of a good cell VM_{good} and a weak cell VM_{weak}; (b) Implementation of the RCRT; (c) Definition of the programmable ratio R in the Read Current Ratio Technique (RCRT); (d) Word Line Pulsing Technique (WLPT) concept; R1 represent a symmetric and R2, R3 - asymmetric defects.

programmable analog levels to the WWTM circuit.

The detected weak cells can be substituted by the available redundant cells or such a chip can be screened out as defective. Moreover, the test results may provide a better insight into the nature and the severity of cell defects and help in debugging. In this paper we present two digitally programmable DFT techniques for SRAM cell stability test utilizing partially discharged floating bit lines. They offer selectable test stress settings and exceed the DRT in defect coverage and detection range.

II. Stability Test Concept

Stability test concept is shown in Fig. 2(a). Solid and dashed lines show the Voltage Transfer Characteristics (VTCs, a.k.a. "butterfly curves") of a good and a weak SRAM cell respectively. In most cases a weak cell is likely to have asymmetrical VTCs due to the presence of V_{TH} and W_{eff}/L_{eff} mismatches and non-catastrophic resistive defects. This results in the shift of the meta-stability point of the cell from Z_{good} to Z_{weak}. If node B is driven below the meta-stable point, then the cell will flip its state. Suppose that we have means to drive node B from V_{node_B} to V_{TEST} such that $VM_{good}<V_{TEST}<VM_{weak}$. Upon removal of the test stress V_{TEST}, node B_{good} of the good cell will return to its previous state. This situation is similar to a non-destructive read operation with incompletely precharged bit lines. Whereas, node B_{weak} of the weak cell will flip to the opposite state (destructive read) as shown by the black and white arrows in Fig. 2(a). A subsequent read operation will reveal the weak cell. By programming the value of V_{TEST}, one can set a past/fail criterion and thus, test for a given degree of cell weakness. All the cells, which flip at the node voltage above V_{TEST} are deemed "weak" (i.e., within the Stability Faults oval in Fig 1(c) top)). The rest of the cells are assumed to have acceptable stability. When $V_{TEST}\rightarrow VM_{good}$, cells with marginal stability may flip causing excessive yield loss. Whereas, when $V_{TEST}\rightarrow VM_{weak}$, defective cells may escape the stability test.

III. Read Current Ratio Technique (RCRT)

The first proposed active programmable technique for SRAM cell stability test is using a set of n SRAM cells in a given column (Fig 2(b)). Either existing cells in the column or external cells can be utilized for this purpose. Let R be the ratio of cells having state "0" to the total number of cells in a set of n cells (Fig. 2(c)). Initially, BL and BLB are precharged to V_{DD}. By writing ratio R into the n cells and simultaneously accessing them with a short word line pulse, one can discharge BL and BLB to the required potentials (Fig. 4(b)). For instance, if the number of cells carrying "0"s and "1"s is equal ($R = 0.5$ in Fig. 2(c)), then, provided the n cells have the same read current, V_{BLB}=V_{BL}. For $R > 0.5$, V_{BL} will be below V_{BLB}, whereas for $R < 0.5$, $V_{BL}>V_{BLB}$. Now, if we access the CUT after BL and BLB have been partially discharged, we can reduce V_{node_A} or V_{node_B} to the desired V_{TEST} value and test the stability of the CUT.

To verify the proposed RCRT, a test chip comprising an asynchronous SRAM and the RCRT implementation has been designed and fabricated in CMOS 0.18μm technology (Fig. 3). Columns (1) include $n = 9$ cells to form ratio R as well as the cells with a variable supply voltage V_{DD_WEAK} to imitate weak cells (10). In the test mode, block (5) issues a short pulse EN_n to the switch gates (9), which simultaneously enables word lines $wl\_1 - wl\_n$ (Fig. 2(b)) of the n cells forming ratio R. The pulse width of $wl_1 - wl_n$ pulse is controlled by a delay element in a one-shot and is proportional to its control voltage $V_{DD\_EN\_ALL}$ (Fig. 4(a), right axis). Depending on the applied pulse width of $wl\_1 - wl\_n$ pulse and on the chosen ratio R, the resulting bit line voltage can can be set in a wide range (Fig. 4(a)left axis, and Fig. 4(b)).

Test sequence starts with writing ratio R in the n cells. Next, the $wl\_1 - wl\_n$ pulse is applied to the n cells and the bit lines are partially discharged. Then, the word line of the CUT (wl_*CUT*) is enabled. If the metastable point of the CUT is above the applied V_{TEST}, such a cell will flip and will be detected by a subsequent read operation. If

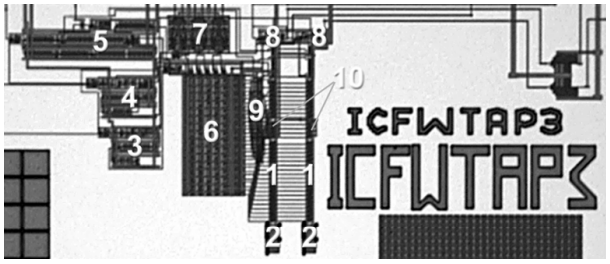
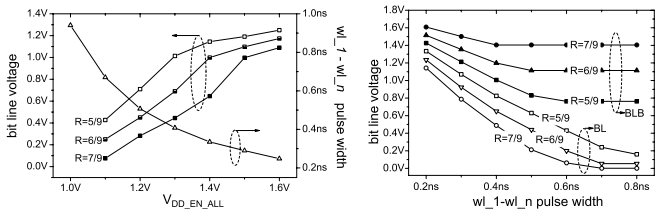


Fig. 3. Microphotograph of the RCRT test chip. (1) SRAM columns with $n = 9$ cells to form ratio R, (2) extra 200fF capacitors to imitate larger C_{BL}, (3-5) read, write and test timing, (6) address decoder, (7) address transition detector, (8) I/O, (9) test enable, (10) cells with a variable supply voltage V_{DD_WEAK} to imitate weak cells.



(a) (b)

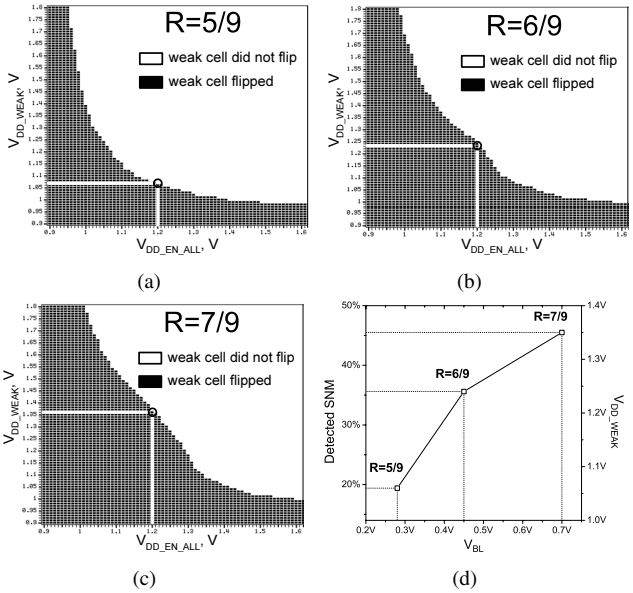
Fig. 4. (a) Bit line voltage and the pulse width of $wl\_1 - wl\_n$ pulse as a function of $V_{DD\_EN\_ALL}$; (b) Bit line voltage as a function of the pulse width of $wl\_1 - wl\_n$ pulse for R=5/9, R=6/9 and R=7/9.

the CUT has flipped, it is marked as defective. Else, the test sequence is repeated with an inverted R to cover possible defects impacting the stability of the other data node of the CUT.

The measurement results showing the detection capability of the RCRT are presented in Fig. 5. Using Agilent 93000 SOC series tester, for each ratio R we swept V_{DD_WEAK} and $V_{DD\_EN\_ALL}$. After applying the test sequence described above, we registered whether the CUT had flipped. The pass/fail (white/black squares) results for R=5/9, R=6/9 and R=7/9 are presented as Shmoo plots in Fig. 5(a)-5(c) respectively. Lower V_{DD_WEAK} values correspond to the smaller SNM of the CUT. Lower $V_{DD\_EN\_ALL}$ values correspond to a wider $wl\_1 - wl\_n$ pulse, which translates into deeper bit line discharge (Fig. 4(b)). An example of the RCRT detection capability for $V_{DD\_EN\_ALL}$ fixed at 1.2V (that corresponds to $wl\_1 - wl\_n$ pulse width of 500ps) is shown in Fig. 5(d). Programming ratio R from $5/9$ to $7/9$ changes the pass/fail threshold from 18% to 46% of the nominal cell SNM. A narrower $wl\_1 - wl\_n$ pulse can further enhance the sensitivity of the RCRT, while using a larger n in ratio R will provide finer programmable step size of V_{TEST} settings.

IV. Word Line Pulsing Technique (WLPT)

Similarly to the RCRT described above, the Word Line Pulsing Technique (WLPT) is also based on the application of partially discharged bit line voltage to the CUT [9]. However, only one existing or external SRAM Reference Cell (REF) per column is used to obtain the reduced voltage value on the bit line. The concept of the WLPT is shown in Fig. 2(d). The WLPT is based on the realization that a precharged bit line BL coupled through the access transistor $Q12$ to the node B



(a) (b) (c) (d)

Fig. 5. (a-c) Shmoo plots for ratios $R = 5/9$, $R = 6/9$ and $R = 7/9$, circled dots represent the detection threshold for $V_{DD\_EN\_ALL} = 1.2V$; (d) Detected SNM as a function of applied V_{BL} (@$V_{DD\_EN\_ALL} = 1.2V$ providing $wl\_1 - wl\_n$ pulse width ≃500ps).

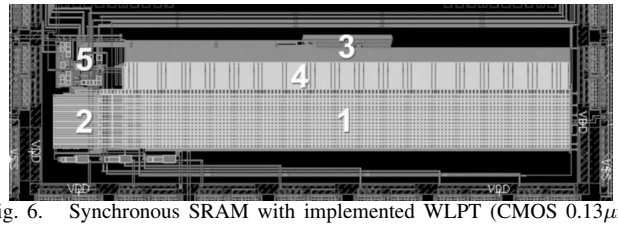


Fig. 6. Synchronous SRAM with implemented WLPT (CMOS 0.13μm): 8Kb SRAM array (256 rows and 32 columns)(1), I/O circuitry (2), address decoder (3) and post-decoder (4), self-timed read/write loop and the Word Line Pulsing Technique (WLPT) timing circuitry (5).

of the REF carrying a "0" (Fig. 2(d)) is gradually discharged by the read current of the REF (I_{read}). The discharge rate of the bit line ΔV_{BL} is a function of the *total duration* that the word line of the REF (WL_REF) has been enabled and can be expressed as $\Delta V_{BL} = (I_{read} * t_{WL\_REF\_pw})/C_{BL}$, where $t_{WL\_REF\_pw}$ is the pulse width of WL_REF and C_{BL} is the bit line capacitance. The total required duration of WL_REF enabled state can be applied as one pulse or broken down into multiple short activation pulses as used in the WLPT. Step-wise discharge of the bit line after each application of a WL_REF pulse is shown in Fig. 7(a). Larger C_{BL} values correspond to finer decrements of V_{BL}.

Stability test mode starts with writing the opposite data backgrounds into the REF and the CUT. After enabling WL_REF for a number of times and reaching the desired V_{BL}, WL_CUT (Fig. 2(d)) is enabled. The access transistor $Q6$ will pass the reduced V_{BL} onto node B of the CUT. The overwrite condition for node B is ensured if we can pull node B below the switching threshold of the inverter formed by transistors $Q1$ and $Q3$. If the effective pull-up drive of node B is weakened by the defect resistance $(R1 + R3)$, the overwrite condition is met earlier and the weak cell is

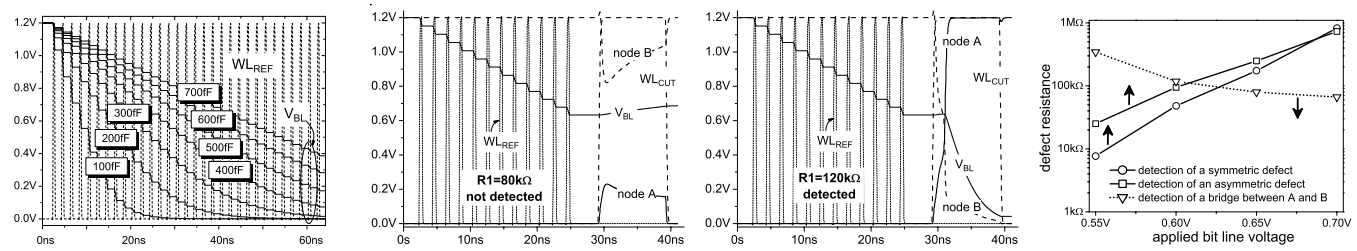


Fig. 7. (a) Bit line discharge as a function of the number of WL_{REF} pulses and C_{BL}; WLPT detection example: (b) R1=80kΩ (not detected) and (c) R1=120kΩ (detected)($C_{BL} = 400fF$); (d) Detection capabilities of the WLPT ($C_{BL} = 400fF$). Arrows show the ranges of detected defect resistance. All simulation results are for CMOS 0.13μm and $t_{WL\_REF\_pw}$=410ps.

overwritten, whereas a good cell with negligible $(R1 + R3)$ will withstand the same stress. Next, the CUT is read back and its data is compared with the previously stored value. If the new read-out is different, the cell is deemed weak. Otherwise, the test sequence is repeated with an inverted background data written into the REF and the CUT.

The WLPT has been implemented in a 8Kb synchronous SRAM test chip (Fig. 6) with a minor modification of the post-decoder (4). Fig. 7(b) and Fig. 7(c) show a detection example of a symmetric defect R1. R1=80kΩ is not detected since node A and node B do not reach the metastable point, whereas R1=120kΩ shifts the metastable point high enough to be detected. The resolving capability (gain-bandwidth product) of an SRAM cell in the metastable region is proportional to the SNM [10]. Therefore, if the SNM of the cell is higher, the cell will have a higher immunity against the metastability and recover from the test disturbance quicker. However, if the stability of the cell is compromised by an excessive process shift or a non-catastrophic defect, such cell will stay in the metastability region longer. Thus, to detect such boundary cases, the duration of the WL_CUT pulse should be sufficient to allow for the extended metastability window. An external larger reference cell and/or a cell with smaller I_{read} as well as connecting all local bit lines in a column together may be used to further refine ΔV_{BL} precision.

The WLPT surpasses the DRT in the detected defect resistance range (Fig. 7(d)). For instance, by applying V_{BL}=0.55V to the CUT, the WLPT can detect R3>20kΩ, whereas the DRT can only detect R3>60MΩ at 150^{o}C. In addition to the defects in the pull-up path (R1, R2 and R3), the WLPT can successfully detect other defects causing poor cell stability, which can be modelled as a resistive bride between node A and node B (R_{A_B})(dotted line in Fig. 7(d)). R_{A_B} can be used as an SRAM cell stability fault model [2]. Reduction of R_{A_B} causes the gain reduction of the two inverters comprising an SRAM cell and proportionally reduces the cell's SNM. The WLPT is demonstrated to detect R_{A_B}<300kΩ (at V_{BL}=0.55V) corresponding to the detected SNM degradation of $\simeq$20% and more [2].

V. Conclusions

We introduced two novel digitally programmable techniques for SRAM cell stability test. Compared to the Data Retention Test, the proposed techniques offer superior defect coverage, reduced test time and no high-temperature requirements. Programmability of the pass/fail threshold facilitates tracking of the process changes and/or quality requirements. That allows to strike a balance between the number of test escapees and the yield loss incurred due to excessively stringent testing. Both the proposed techniques offer flexible implementation options and small area overhead.

Normally, implementation involves minimal modifications of the address decoder and BIST without extra hardware in SRAM array. More capacitive bit lines allow for setting a finer step size of the test stress applied to the CUT, which improves the resolution of the proposed techniques. The obtained test information enables to distinguish symmetric and asymmetric defects in the pull-up path of SRAM cells as well as to assess the degree of the cell stability degradation.

VI. Acknowledgment

Authors would like to thank Patrick van de Steeg (Philips Semiconductors), Roelof Salters, Clemens Wouters, Rutger van Veen and Bram Kruseman (Philips Research Labs) for fruitful discussions and support with measurements.

References

[1] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid State Circuits*, vol. 36, pp. 658–665, Apr. 2001.

[2] A. Pavlov, M. Sachdev, and J. P. de Gyvez, "An SRAM weak cell fault model and a DFT technique with a programmable detection threshold," in *Proc. IEEE International Test Conference (ITC)*, Oct. 2004, pp. 1106–1115.

[3] International technology roadmap for semiconductors - 2004 update.

[4] P. Stolk, H. Tuinhout, *et al.*, "CMOS device optimization for mixed-signal technologies," in *IEEE International Electron Devices Meeting IEDM Technical Digest*, Oct. 2001, pp. 10.2.1–10.2.4.

[5] A. Meixner and J. Banik, "Weak write test mode: An SRAM cell stability design for test technique," in *Proc. IEEE International Test Conference (ITC)*, Nov. 1997, pp. 1043–1052.

[6] T. W. Liston and L. N.Herr, "Method and apparatus for soft defect detection in a memory," U.S. Patent 6 590 818, July 8, 2003.

[7] M. Mehalel, "Short write test mode for testing static memory cells," U.S. Patent 6 256 241 B1, July 3, 2001.

[8] E. B. Selvin, A. R. Farhang, and D. A. Guddat, "Programmable weak write test mode," U.S. Patent 6 778 450 B2, Aug. 17, 2004.

[9] J. P. de Gyvez, M. Azimane, and A. Pavlov, "SRAM test method and SRAM test arrangement," Patent pending.

[10] L.-S. Kim and P. W. Dutton, "Metastability of CMOS latch/flip-flop," *IEEE J. Solid-State Circuits*, vol. 25, pp. 942–951, Aug. 1990.