

A Dense 45nm Half-differential SRAM with Lower Minimum Operating Voltage

Gregory Chen, Michael Wieckowski, David Blaauw, Dennis Sylvester

University of Michigan, Ann Arbor, Michigan
 {grkchen, wieckows, blaauw, dmcs}@umich.edu

Abstract- We present a 45nm half-differential 6T SRAM (HD-SRAM) with differential write and single-ended read, enabling asymmetric sizing and V_{TH} selection. The HD-SRAM bitcell uses SRAM physical design rules to achieve the same area as a commercial differential 6T SRAM (D-SRAM). We record measurements from 80 32kb SRAM arrays. HD-SRAM is 18% lower energy and 14% lower leakage than D-SRAM. It has a 72mV-lower V_{MIN} , demonstrating higher stability.

I. INTRODUCTION

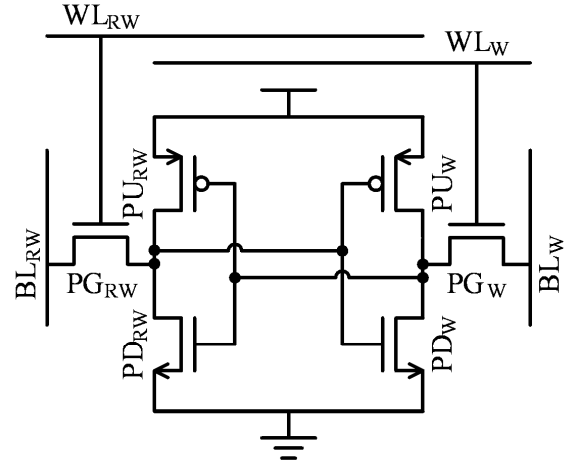
Process variations such as random dopant fluctuation and line edge roughness degrade SRAM operating margins [1]. Since designs commonly have large SRAMs, each bitcell must be extremely robust to achieve high chip yield. In differential 6T SRAM (D-SRAM), read stability is improved by making the pull down (PD) device strong relative to the pass gate device (PG). This reduces the probability of read upset or destructive read failures when bitcells are subject to process variation. Write stability is improved by making PG strong relative to the pull up device (PU). SRAM designs are commonly read-stability limited.

Many SRAM designs achieve higher read stability by increasing PD width and PG length. In addition, PG typically has a higher threshold voltage (V_{TH}) than PD. However, making the device dimensions larger increases bitcell area. Also, increasing PG L and V_{TH} reduces performance and degrades write margins. The two-sided read and write criteria create an upper bound on the overall stability margin. 8T bitcells separate read and write circuitry to increase stability at the expense of area and leakage [2]. The proposed half-differential 6T SRAM (HD-SRAM) improves voltage scalability and operating margin with no increase in bitcell size or leakage.

II. HALF-DIFFERENTIAL SRAM METHOD

A. Operation, Sizing and V_{TH} Selection

HD-SRAM performs a differential write access in the same manner as D-SRAM, but only reads the bitcell from one side (Fig. 1). This enables asymmetric sizing and V_{TH} -selection optimizations to improve stability margins. During a write operation, both wordlines (WLs) are asserted, both PGs turn on, and the differential value on the bitlines (BLs) overwrites the cell value. During a read operation, the read-and-write WL (WR_{RW}) is asserted, turning on the read-and-write PG (PG_{RW}). This device selectively discharges its associated BL (BL_{RW}) based on the bitcell's stored value.



Device	W, nm	L, nm	V_{TH}
PD _{RW}	240	55	LVT
PD _W	100	95	HVT
PU _{RW}	60	55	SVT
PU _W	60	95	SVT
PG _{RW}	150	105	HVT
PG _W	100	65	HVT

Figure 1. HD-SRAM operates with differential write and single-ended read, enabling asymmetric sizing and V_{TH} selection for higher robustness.

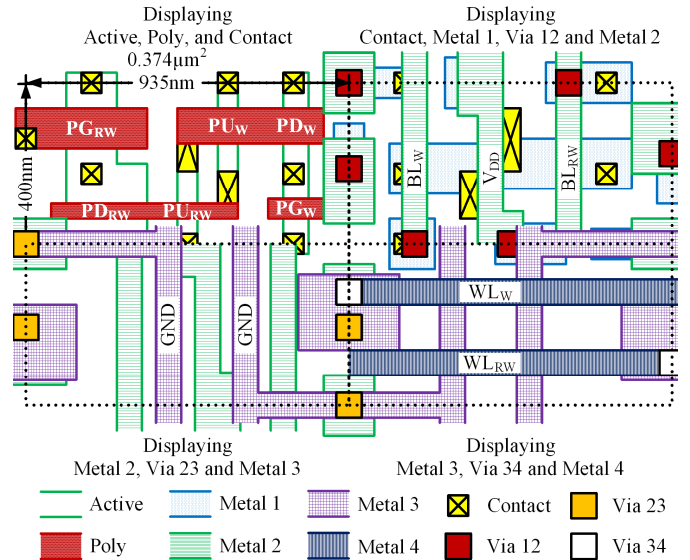


Figure 2. HD-SRAM is the same size as a commercial differential 6T design (D-SRAM). Both designs exceed logic design rules for higher density.

Asymmetric sizing and V_{TH} -selection optimizations increase bitcell stability without increasing area or energy. For the single-ended read, the write-only pull down device (PD_W) does not strongly impact read stability, so we reduce its width to minimum size. This significantly reduces the bitcell area because in D-SRAM, the PDs are large to enhance read stability. We apply the resulting area savings to increase the read-and-write side PD (PD_{RW}) width and PG (PG_{RW}) length, improving read margin. Since the length of PG_{RW} is increased, we can increase the lengths of PD_W and the write-only pull-up device (PU_W) no area penalty. This increases write-one margin and also improves read stability by decreasing the positive feedback between the cross-coupled inverters.

In many D-SRAM designs, two NMOS V_{TH} s are available in the process, a higher V_{TH} for PGs and a lower V_{TH} for PDs. For each HD-SRAM device, we optimally select from the existing V_{TH} s to improve stability margins. The higher NMOS V_{TH} usually reserved for PGs is used for PD_W to help prevent read upsets. Using the low V_{TH} device for the write-only pass gate (PG_W) would further increase write-one margin but decreased the overall simulated robustness and increased leakage. So the higher V_{TH} is selected for this device. Previous asymmetric SRAMs do not provide silicon results and either decrease robustness, increase bitcell area, and/or do not consider the physical design of SRAM [3][4][5].

B. Physical Design

The HD-SRAM bitcell has the same area ($0.374\mu m^2$) as the commercial D-SRAM bitcell in this 45nm process to allow for an accurate comparison. The layout violates logic design rules to achieve higher density, which is typical for commercial SRAM but uncommon in research efforts [3] (Fig. 2). We implemented the design with feedback from the foundry regarding design, lithography, and design for manufacturing (DFM) rules. The two WLs are on Metal 4, grounds are on Metal 3, and the BLs and V_{DD} are on Metal 2. All polysilicon is linear and unidirectional to enable double patterning. Unlike most D-SRAM, PD_W and PG_W are the same width in HD-SRAM, eliminating a notch in the source-drain region and improving DFM.

C. Simulated Results

HD-SRAM achieves higher robustness than D-SRAM, even when peripheral assist circuits and optimal technology selection are applied only to D-SRAM. HD-SRAM has an 85-mV higher simulated static noise margin (SNM) than D-SRAM at the nominal V_{DD} of 1.1V (Fig. 3a). The HD-SRAM SNM remains higher as V_{DD} scales to below 500mV (Fig. 3b).

Since SRAM is typically read-stability limited at nominal V_{DD} , one read assist technique reduces the WL voltage (V_{WL}) to increase read margin [6]. As a measure of robustness, we simulate the maximum V_{TH} variation that a typical bitcell can tolerate without functional failure for read, write, and hold operations. We simulate the designs in SPICE using importance sampling and normalize the robustness to a typical 45nm distribution of V_{TH} with $\sigma=40mV$ [7]. As D-SRAM V_{WL}

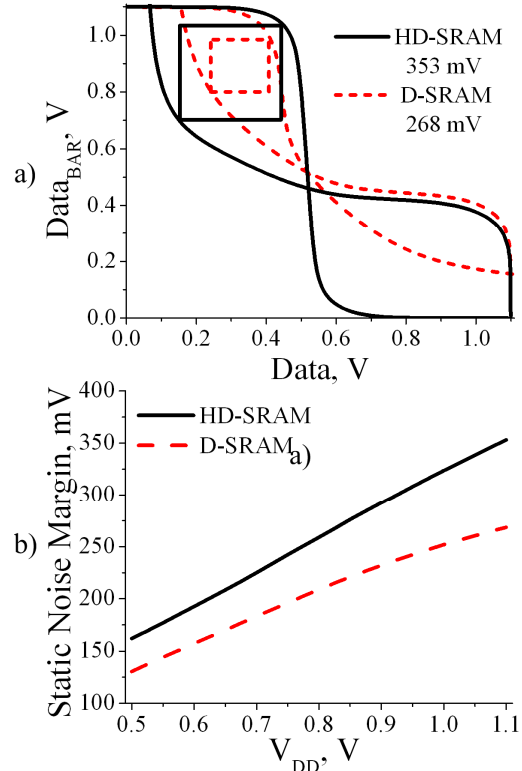


Figure 3. HD-SRAM has an 85mV-higher simulated SNM than D-SRAM at nominal V_{DD} . SNM remains higher as V_{DD} scales below 500mV.

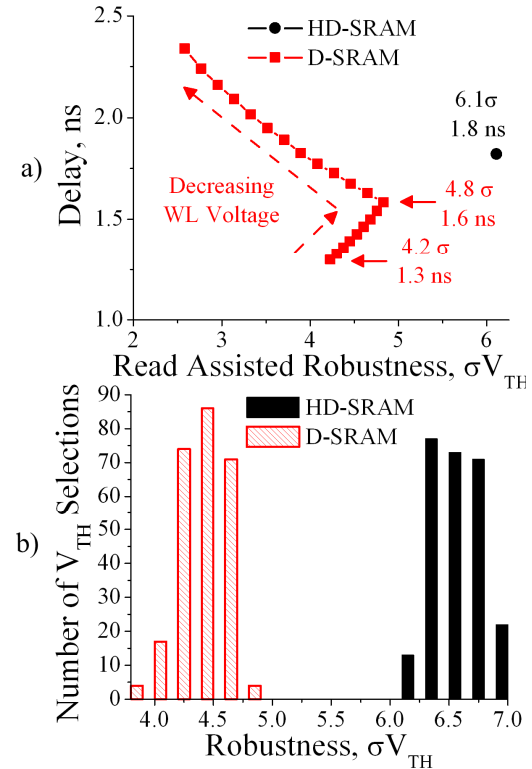


Figure 4. D-SRAM robustness improves with assist techniques such as WL voltage selection (a) and technology V_{TH} selection (b). However, neither of these techniques achieves as high robustness as HD-SRAM.

decreases from 1.1V to 1.02V, read-stability and total robustness increase from 4.2□ to 4.8□ (Fig. 4a). However, as V_{WL} further decreases, write margin degrades overall robustness and latency becomes prohibitive. Separate voltages can be used for write and read, but this requires pre-decoding and additional complexity. HD-SRAM without read assistance is more robust than D-SRAM at any V_{WL} . HD-SRAM robustness further improves with read assistance.

The optimal selection of technology parameters, such as V_{TH} , also improves robustness. In typical SRAM processes, these parameters are carefully tuned to optimize the design. However, the nominal V_{TH} selections may trade off robustness for improved performance. We simulate bitcell robustness in SPICE using importance sampling for theoretical selections of technology parameters, with reasonable selections of PD, PG and PU V_{TH} s. The maximum D-SRAM robustness of 4.8□ is achieved by reducing PD V_{TH} and increasing PG V_{TH} (Fig. 4b). This robustness is lower than both the nominal and maximum HD-SRAM robustness of 6.1□ of 7.0□, respectively.

III. MEASUREMENT RESULTS

A. Test Chips

We fabricated test chips including 32kb banks of HD-SRAM and commercial D-SRAM in a 45nm CMOS process with 1.1V nominal V_{DD} (Figs. 5 and 6). Each bank uses identical address decoders, WL and BL drivers, and sense amplifiers (SAs). HD-SRAM adds gating logic and an additional WL driver to support two WLs per row, slightly decreasing array efficiency. We tie one HD-SRAM SA input to a reference voltage to accommodate single-ended read. The test chips do not include assist circuits, error correction coding (ECC), or redundancy, which could be applied to either design. A BIST performs functionality and performance tests on each design. Functionality is assessed by performing march tests with solid, checkerboard and stripe test patterns.

B. Performance, Power and Leakage

D-SRAM is 15% faster than HD-SRAM. Performance is defined as the speed at which every bitcell in the array is functional. It is read limited and includes WL, bitcell, and BL delays for each design (Fig. 7). In a microprocessor, this delay amortizes over register, interconnect, decoder, sense amplifier and multiplexer delays. HD-SRAM has larger read devices that exhibit less timing sensitivity to process variation, decreasing array latency that is dictated by the slowest cells.

HD-SRAM has 18%-lower access energy than D-SRAM (Fig. 8a). The HD-SRAM write energy is slightly higher because of higher total capacitance on the WLs caused by higher routing capacitance. However, the HD-SRAM read energy is significantly lower since only WL_{RW} switches and capacitance on this WL is lower than the total D-SRAM WL capacitance. Also, in the read-one case, neither BL discharges, whereas one BL always discharges during a D-SRAM read.

HD-SRAM has a 14%-lower leakage power than D-SRAM (Fig. 8b). The leakage improvements result from longer gate lengths selected for PG_{RW} , PD_w , and PU_w . In

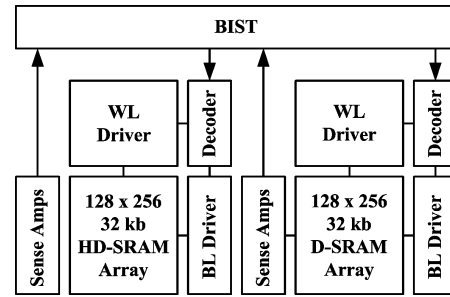
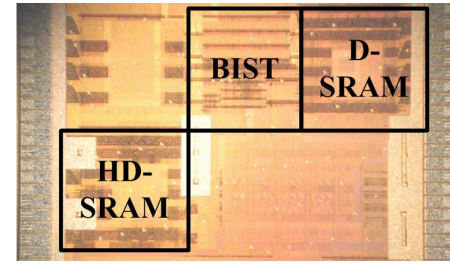


Figure 5. HD-SRAM and D-SRAM arrays use nearly-identical peripheral circuits. A BIST performs march and speed tests on both designs.



	HD-SRAM	D-SRAM
Process	45nm CMOS	
Bitcell Area	0.37 μm^2	
μ R+W Margin	12.1 σ	11.0 σ
μV_{MIN}	639 mV	711 mV
Sim. SNM	353 mV	268 mV
Performance	550 MHz	650 MHz
Energy / bit	43 fJ	53 fJ
Leakage / bit	55 pW	64 pW

Figure 6. Chip micrograph and results summary.

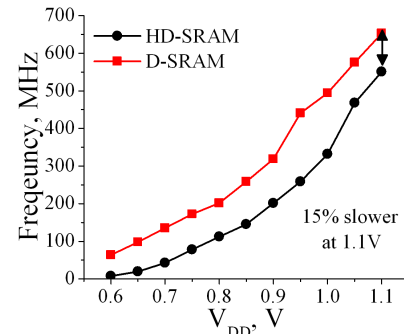


Figure 7. Measured performance results show that the HD-SRAM array is 15% slower than D-SRAM at nominal V_{DD} . Array performance is dictated by the slowest cells and HD-SRAM exhibits less timing variation.

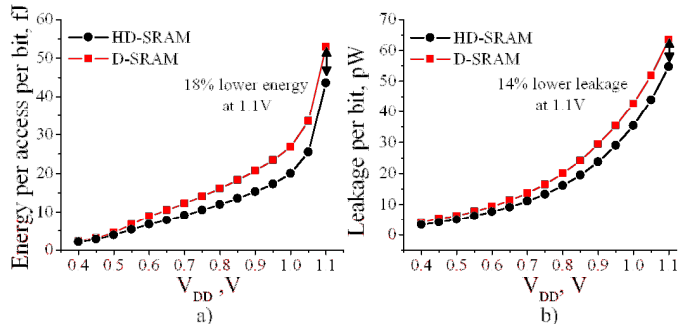


Figure 8. HD-SRAM has an 18%-lower measured access energy and a 14%-lower measured leakage than D-SRAM.

addition, P_{D_W} has a higher V_{TH} than in D-SRAM. 8T SRAM would have significantly higher leakage than D-SRAM, because of additional devices added to the bitcell.

C. Minimum Operating Voltage

We record the minimum operating voltage (V_{MIN}) for 80 test chips. V_{MIN} is defined as the V_{DD} where all bitcells in the SRAM bank are functional. A lower V_{MIN} indicates higher bitcell stability, since the bitcell is more tolerant to the decreased noise margins and the greater effect of process variations at low voltage. Error maps from one chip show that V_{MIN} is 800mV for D-SRAM, with one bitcell failing below this voltage. Meanwhile, every HD-SRAM bitcell functions at 650mV, demonstrating a lower V_{MIN} for this array (Fig. 9).

Across all 80 test chips, the average HD-SRAM V_{MIN} is 639mV, whereas the average D-SRAM V_{MIN} is 72 mV higher at 711mV (Fig. 10). HD-SRAM also exhibits decreased variation in V_{MIN} among test chips because the devices that limit stability are larger and less affected by process variations such as random dopant fluctuation and line edge roughness. Only 4 HD-SRAM arrays have V_{MIN} above 700mV, whereas 35 D-SRAM arrays fail this criterion. This demonstrates higher HD-SRAM yield at a given V_{DD} .

Bitcell failure rates are recorded for all 80 test chips. At nominal V_{DD} , SRAM failures are rare. Therefore, to observe a significant number of errors, V_{WL} is raised by 50mV to aggravate read failures. Since these cells are typically read stability limited, this emphasizes variation and emulates cells at the tails of the process variation distributions. In larger SRAM arrays, it is more likely that these tail bitcells would appear. Under this condition, HD-SRAM has a 100 \times lower failure rate than D-SRAM at nominal V_{DD} (Fig. 11).

ACKNOWLEDGEMENTS

The authors thank STMicroelectronics for fabrication and support of this project.

REFERENCES

- [1] R. Aitken, S. Igdunji, "Worst-Case Design and Margin for Embedded SRAM," *Design, Automation & Test in Europe Conference*, pp.1-6, Apr. 2007.
- [2] L. Chang et al., "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE Journal of Solid-State Circuits*, vol.43, no.4, pp.956-963, Apr. 2008
- [3] N. Azizi, F.N. Najm, A. Moshovos, "Low-leakage Asymmetric-cell SRAM," *IEEE Transactions on VLSI*, vol.11, no.4, pp. 701- 715, Aug. 2003.
- [4] B. S. Gill, C. Papachristou, F.G. Wolff, "A New Asymmetric SRAM Cell to Reduce Soft Errors and Leakage Power in FPGA," *Design, Automation & Test in Europe*, pp.1-6, Apr. 2007.
- [5] K. Kim; J.-J. Kim, C.-T. Chuang, "Asymmetrical SRAM Cells with Enhanced Read and Write Margins," *International Symposium on VLSI Technology, Systems and Applications*, pp.1-2, 23-25 Apr. 2007.
- [6] K. Nii et al., "A 45-nm Bulk CMOS Embedded SRAM with Improved Immunity Against Process and Temperature Variations," *IEEE Journal of Solid-State Circuits*, vol.43, no.1, pp.180-191, Jan. 2008.
- [7] G.K. Chen, D. Blaauw, T. Mudge, D. Sylvester, and N.S. Kim, "Yield-driven Near-threshold SRAM Design," *IEEE/ACM International Conference on Computer-Aided Design*, pp.660-666, 4-8 Nov. 2007.

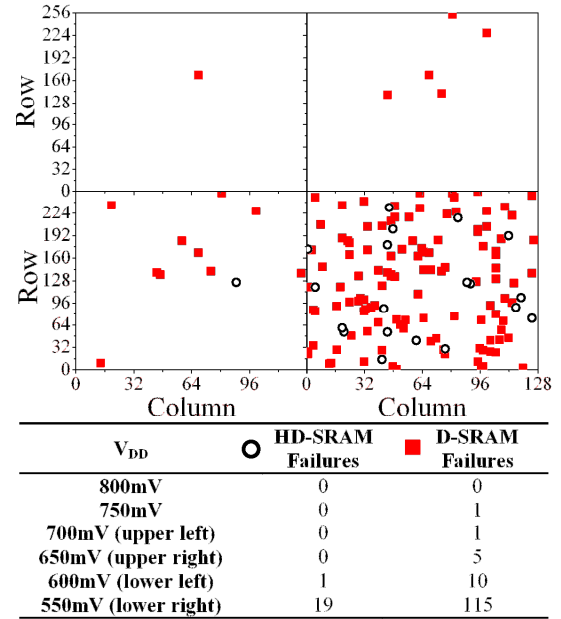


Figure 9. Failure maps show bitcell failure locations as V_{DD} is scaled down. For this test array V_{MIN} is 650mV for HD-SRAM and 800mV for D-SRAM.

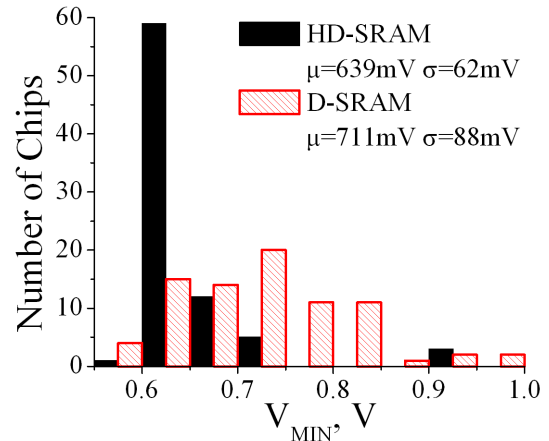


Figure 10. A histogram of measured V_{MIN} for 80 test chips shows that HD-SRAM has a 72mV-lower average V_{MIN} and fewer arrays with high V_{MIN} .

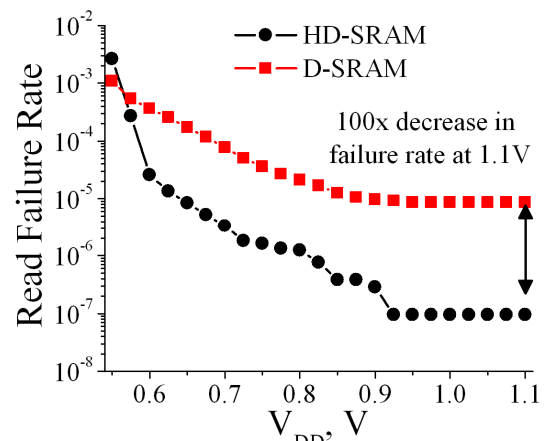


Figure 11. At nominal V_{DD} , HD-SRAM has a 100 \times -lower bitcell failure rate than D-SRAM. Read failures dominate cell stability at nominal V_{DD} and are aggravated in only this plot to observe a significant number of failures.