A 40-nm 256-Kb Sub-10 pJ/Access 8T SRAM with Read Bitline Amplitude Limiting (RBAL) Scheme

[1]Shusuke Yoshimoto, [1]Masaharu Terada, [1]Youhei Umeki, [1]Shunsuke Okumura,
[2]Atsushi Kawasumi, [2]Toshikazu Suzuki, [2]Shinichi Moriwaki, [2]Shinji Miyano,
[1]Hiroshi Kawaguchi, and [1]Masahiko Yoshimoto
[1]Graduate School of System Informatics, Kobe University, Kobe, Japan
[2]Semiconductor Technology Academic Research Center, Yokohama, Japan
yoshipy@cs28.cs.kobe-u.ac.jp

ABSTRACT
This paper presents a novel read-bitline amplitude limiting (RBAL) scheme which suppresses dynamic energy dissipation caused by random variation. In addition, a discharge acceleration (DA) circuit is proposed to decrease delay overhead of RBAL. The proposed scheme improves the active energy dissipation in a read cycle by 22% at the center–center (CC) corner and 25°C. The maximum delay overhead is 32% at the fast–slow (FS) corner and -40°C. The circuits have been implemented using the 40-nm bulk CMOS process. The implemented 256-Kb 8T SRAM works fine with energy dissipation of sub-10 pJ / access from 0.5–0.7 V.

Categories and Subject Descriptors
B.3.1 [**Memory Structures**]: Semiconductor Memories – *Static Memory (SRAM)*

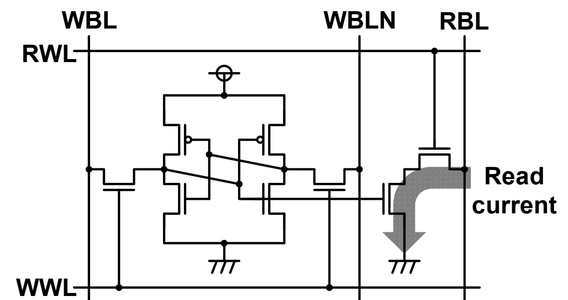
General Terms
Design

Keywords
8T SRAM, low voltage, low energy, read bitline limiter, discharge accelerator

1. INTRODUCTION
The minimum feature size in transistors continues to decrease with the advance of process technology. Process scaling realizes higher density and lower cost. In the deep sub-micron era, the threshold voltage (*Vt*) deviation in transistors is, however, increasing to more than 100 mV as $3\sigma_{Vt}$ [1–3]. Consequently, designing a 6T SRAM cell has become increasingly difficult: both read and write margins must be considered [4–7]. The 8T SRAM cell presented in Fig. 1(a) was proposed to eliminate read failures caused by the dedicated read port. Therefore, in the 8T cell, only the write margin must be considered, which can make a layout smaller and less expensive than the 6T cell in future processes [8].

Recent reports have described the effectiveness of low-voltage operation for energy reduction in SRAM degrades at the scaled process [9–10]. Figure 1(b) presents a simulated histogram of the read current in 8T SRAM cell at 0.5 V. The read current of the slowest cell in five thousand Monte-Carlo simulations was 0.016 times slower than the nominal cell. SRAM designers configure a pulse width of the read wordline (RWL) by the slowest cell and the sense-amp offset. Figure 2 depicts simulated waveforms of the read bitline (RBL) including random variation. In most cases, RBLs are fully discharged, although the slowest cell is read out, which increases the dynamic energy by 82% compared to the nominal simulation without random variation [10].

Realizing low-energy operation at low supply voltage demands reduction of the energy increase caused by random variation. An earlier proposed bitline amplitude limiter (BAL) suppresses the energy increase shown in Fig. 3(a) [10], limiting



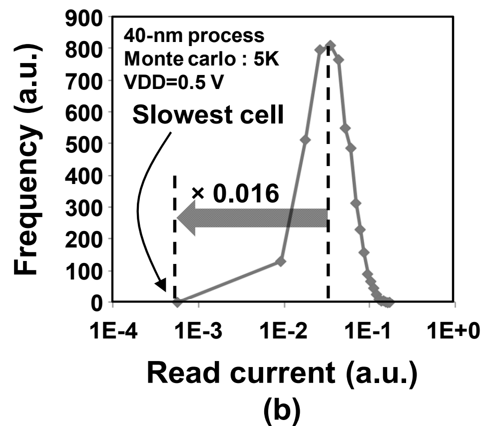


Figure 1. Random variation in read ports of (a) single-ended 8T SRAM cells and (b) histgram of read current.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ISLPED'12, July 30–August 1, 2012, Redondo Beach, California, USA.
Copyright 2012 ACM 978-1-4503-1249-3/12/07...$10.00.

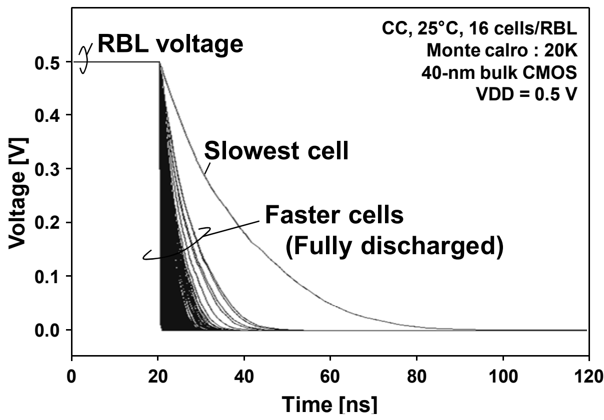


Figure 2. Waveforms of read bitline in single-ended 8T cell at 0.5 V.

the swing of bitlines because half-swing bitlines make the virtual ground of memory cells floating, and reducing the dynamic and leakage power by 26%. However, BAL presents the disadvantage of disturb-margin degradation. The 6T cell uses differential bitlines to read. Consequently, BAL consists of two series of NMOSes, which degrades the cell stability at low voltage.

As described in this paper, we proposed a read bitline amplitude limiter (RBAL) using 8T SRAM cell and a novel discharge acceleration (DA) scheme, presented in Fig. 3(b). RBAL is applied to the disturb-free single-ended read port of the 8T cell. Therefore, disturb-margin degradation and RBAL consists of only one NMOS. The DA scheme is proposed to improve the delay overhead of RBAL.

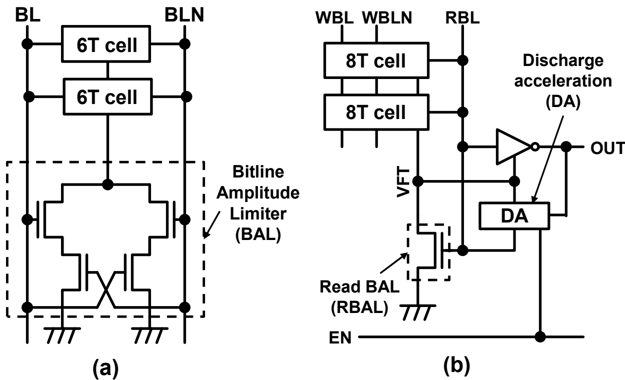


Figure 3. Diagrams of (a) Bitline amplitude limiter (BAL) with 6T cell and (b) read bitline amplitude limiter (RBAL) and distcharge acceleration circuit with 8T cell.

2. READ BITLINE AMPLITUDE LIMITER AND DISCHARGE ACCELERATION SCHEME

Figure 4 portrays an RBAL and novel DA scheme, which consist of only three NMOSes. The gate and drain of RBAL are connected respectively to RBL and virtual footer (VFT). In a 1-read operation, the RBLs are pulled down by the activated read ports. Figure 5 presents the waveforms of RBL with RBAL and without the DA scheme. Although the slowest cell is discharging the RBL, the other RBLs are pulled down faster. However, the discharge is stopped by the threshold voltage of RBAL. The

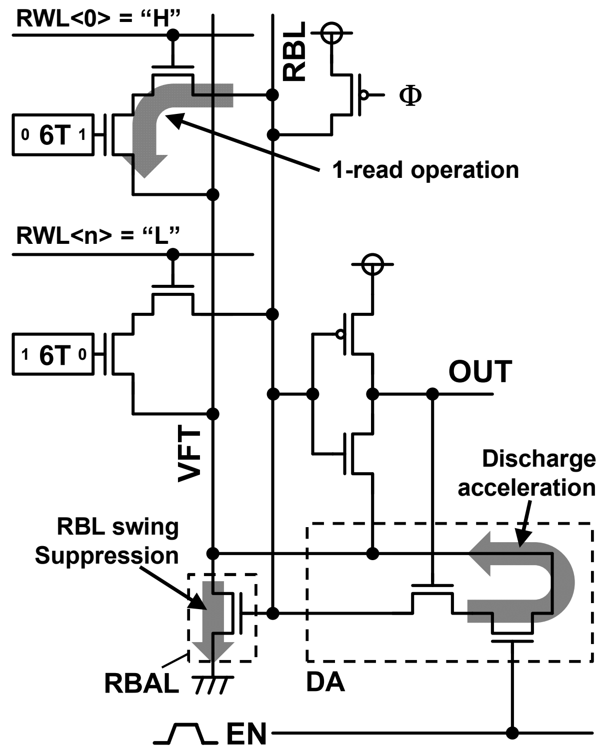


Figure 4. Circuit diagram of RBAL and discharge acceleration (DA) scheme.

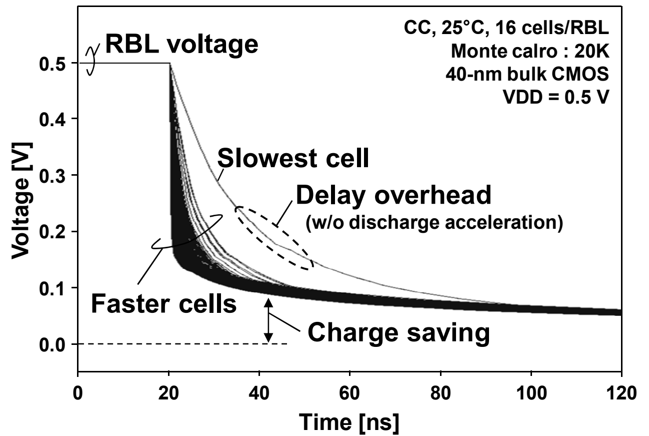


Figure 5. Waveforms of read bitline in proposed 8T cell with read bitline limiter.

decrease of average swing in RBL improves the dynamic energy. The single-ended 8T cell generally uses an inverter as a sense amp. Consequently, there is delay overhead when the RBL ranges under middle voltage. To prevent the delay degradation, the novel DA scheme is proposed. The DA terminals are connected to RBAL, VFT, output of sense-amp (= inverter) and enable signal (EN). When the output voltage is over the threshold voltage of the NMOS and the EN signal is activated, the RBL and the VFT are shorted by the two series of NMOSes. The DA scheme improves the read bitline delay presented in Fig. 6.

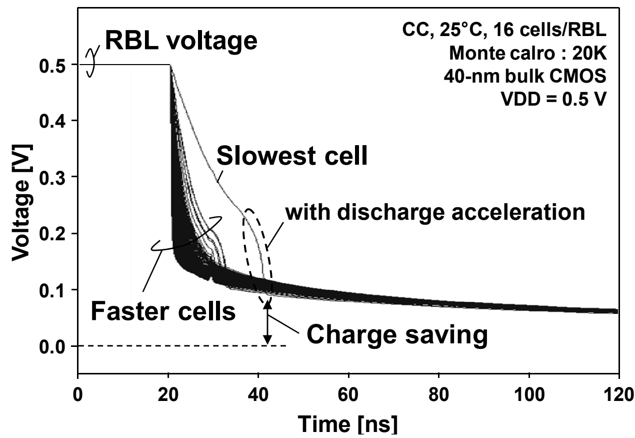


Figure 6. Waveforms of read bitline in proposed 8T cell with read bitline limiter and assist circuit.

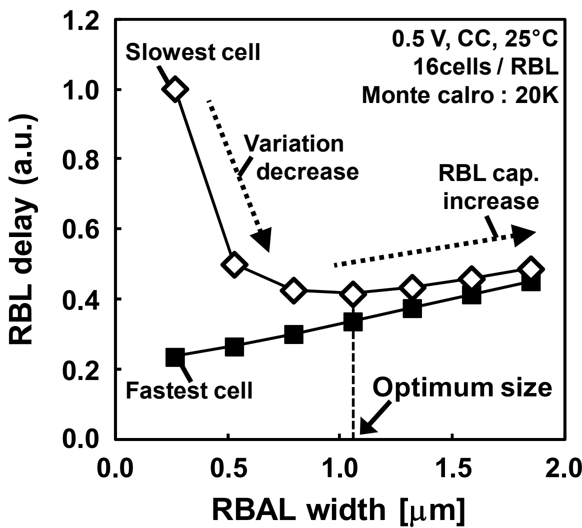


Figure 7. Read bitline (RBL) delay versus the width of the limiter transistor.

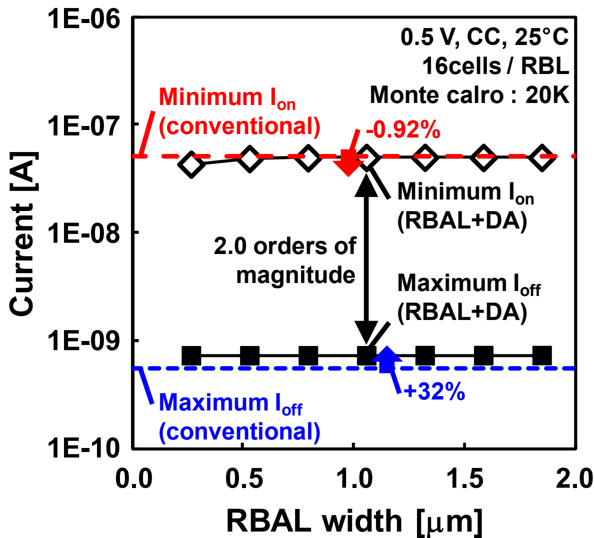


Figure 8. Current margin.

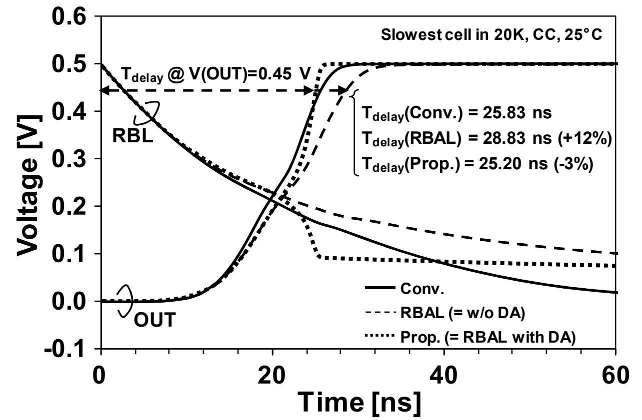


Figure 9. Waveforms of the slowest cell.

Figure 7 portrays RBL delay versus width of the RBAL. Smaller RBAL increases the threshold variation. However, longer RBAL increases the capacitance of RBL. Herein, the width of RBAL is optimized by the RBL delay.

Figure 8 presents the minimum drain current (I_{on}) and maximum leakage current (I_{off}) of the read port in the 8T cell. The number of 8T cells in a RBL is 16. The red line and blue line respectively indicate the minimum I_{on} and I_{off} in a conventional 8T cell. The proposed scheme with RBAL and DA decreases the minimum I_{on} by 0.92% and increases the maximum I_{off} by 32%, however, the I_{on} is two orders of magnitude larger than I_{off}.

Figure 9 shows waveforms of the slowest cell in conventional, RBAL without DA and RBAL with DA when 20-K Monte-Carlo analyses are executed at CC, 25°C. In this paper, T_{delay} is defined by the time to which V(OUT) rises to 0.45 V at supply voltage of 0.5 V. The T_{delay} with only RBAL is increased by 12%. However, the T_{delay} with RBAL and DA is decreased by 3%. Table 1 and 2 respectively show T_{delay} and active energy (E_{active}) improvement at five process corners and temperatures of three kinds. T_{delay} at the FS corner are increased by 14%-32%, however, T_{delay} at SS corner are improved by 2%–5%. E_{active} are improved by 13%–27% at 0.5 V.

Table 1. T_{DELAY}.

		T_{delay} @ V(OUT) = 0.45 V [ns]			T_{delay} ratio
		Conv.	RBAL	Prop.	Prop./Conv.
SS	-40	1761.30	2008.00	1700.30	0.97
	25	145.62	161.31	138.97	0.95
	125	15.12	16.40	14.75	0.98
SF	-40	518.29	547.16	457.84	0.88
	25	55.01	58.33	49.26	0.90
	125	7.28	7.74	7.74	1.06
CC	-40	175.97	205.83	173.40	0.99
	25	25.83	28.83	25.20	0.98
	125	4.52	4.91	4.91	1.09
FS	-40	53.41	86.61	70.62	1.32
	25	11.35	14.54	13.61	1.20
	125	2.69	3.07	3.07	1.14
FF	-40	16.04	18.91	16.78	1.05
	25	4.51	5.02	5.02	1.11
	125	1.36	1.47	1.47	1.08

Table 2. E_{ACTIVE}.

		E_{active} in a RBL [fJ]			E_{active} ratio
		Conv.	RBAL	Prop.	Prop./Conv.
SS	-40	3.1320	2.3595	2.3398	0.75
	25	3.1536	2.4799	2.4507	0.78
	125	3.2174	2.6706	2.6024	0.81
SF	-40	3.1482	2.3205	2.2916	0.73
	25	3.1997	2.4681	2.4204	0.76
	125	3.2968	2.6722	2.6722	0.81
CC	-40	3.1439	2.3748	2.3439	0.75
	25	3.1822	2.5260	2.4716	0.78
	125	3.2884	2.7589	2.7589	0.84
FS	-40	3.1525	2.4406	2.4035	0.76
	25	3.2028	2.6159	2.5188	0.79
	125	3.4060	2.9409	2.9409	0.86
FF	-40	3.1970	2.4570	2.3821	0.75
	25	3.2668	2.6497	2.6497	0.81
	125	3.4509	2.9605	2.9605	0.86

3. CHIP IMPLEMENTATION AND MEASUREMENT RESULTS

We designed and fabricated a 256-Kb SRAM macro in a 40-nm bulk CMOS process, as presented in Fig. 10. Table 3 shows the test chip configuration: The macro size is 0.566 mm × 0.976 mm, the 256-Kb SRAM macro consists of 128 rows × 128 columns × 16 banks, the 8T cell size is 0.706 µm based on logic rule, and the cell density is 463 Kb / mm^2. Figure 11 portrays 16-Kb sub block with the proposed circuit: RBAL and DA. The sub block consists of 16 local cell array (8 columns × 128 rows). Figure 12 portrays a local cell array, which includes the local read circuit, RBAL, DA and low-swing bitline driver (LSBD) which achieves low-power write-back function [11]. At a write cycle, a RWL is activated and an LSBD drives a pair of WBL/WBLN in a half-selected column according to the readout data. Figure 13 portrays the circuit details. The global RBL (GRBL) is not activated during the write cycle to suppress the energy dissipation. The RBAL and DA transfer the readout datum to the LSBD at the write cycle and to GRBL at the read cycle. The LSBDs are activated only in half-selected columns according to the column line enable (CLE) and drive enable bar (DRN). In write target columns, CMOS write drivers activate the write bitlines so the write margin is not degraded. Figure 13 shows the measured energy dissipation per cycle (R:W = 50:50) in the minimum operation voltage (VDD_{min}) of 10 MHz, 20 MHz and 100 MHz at room temperature (RT). The measurement results are investigated using 16 sample chips. Consequently, the energy consumption is less than 10 pJ / access at 0.5–0.7 V.

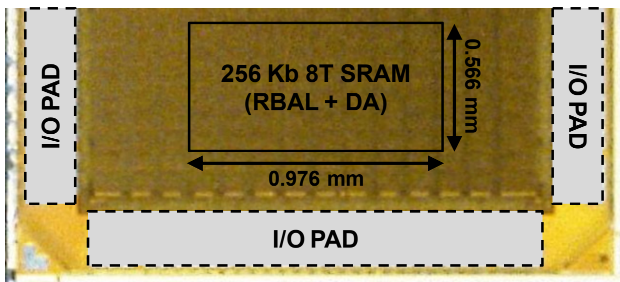


Figure 10. Die photograph of the test chip.

Table 3. Test chip configuration.

Technology	40 nm bulk CMOS
Macro size	0.566 mm x 0.976 mm
Macro configuration	256 Kb (16Kb x 2 x 8), 16 bits/word
Cell size	0.706 µm (logic rule)
# of cells/BL	16 (RBL), 128 (WBL)
Density	463 Kb/mm^2

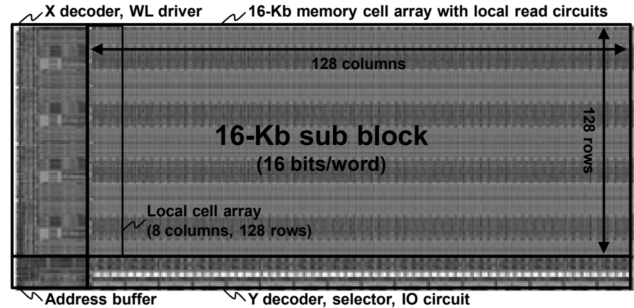


Figure 11. 16-Kb sub block with the proposed circuitry.

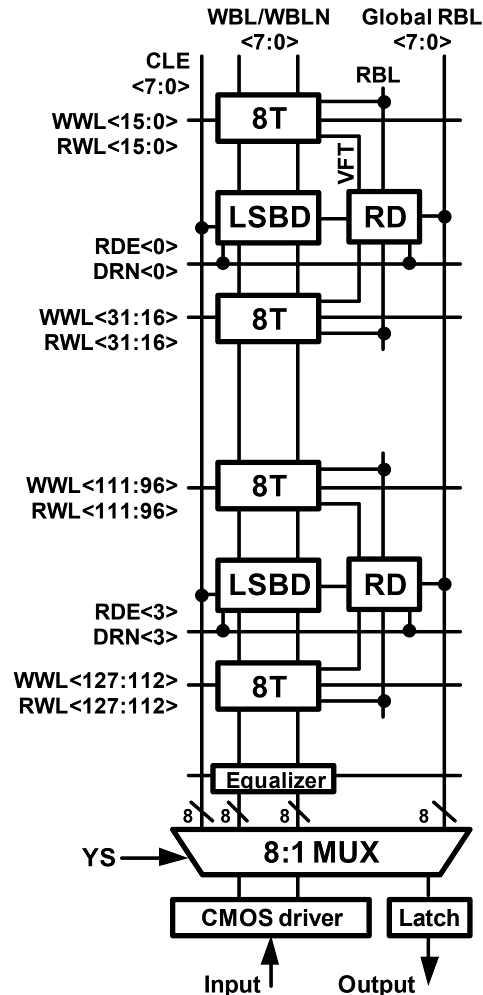


Figure 12. Diagram of local cell array.

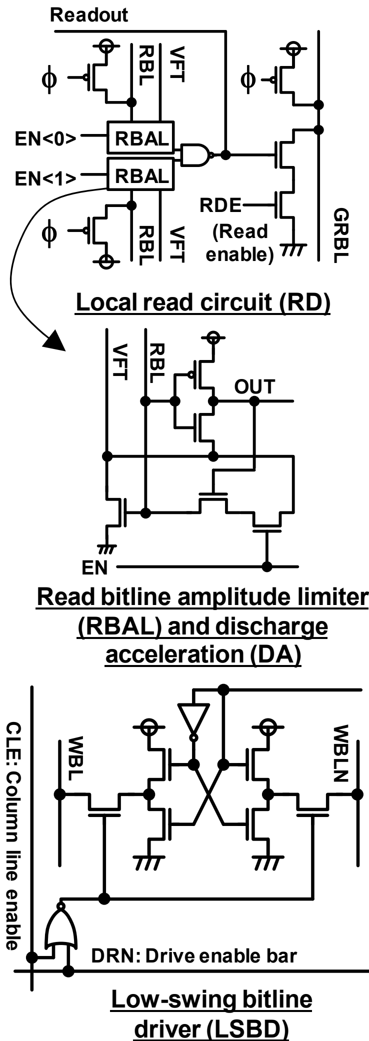


Figure 13. Diagram of the proposed circuits and low-swing bitline driver (LSBD) to prevent the half-select problem [11].

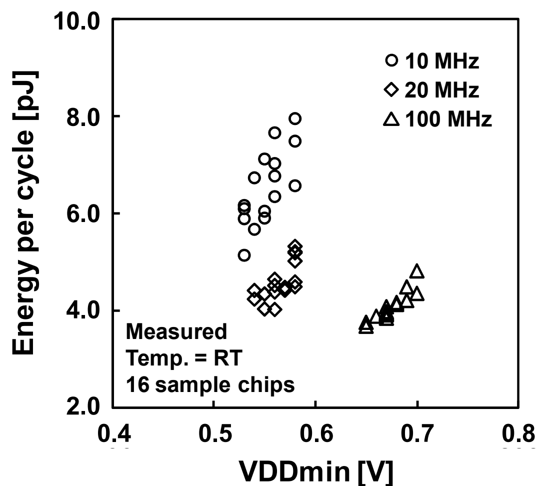


Figure 14. Measured energy dissipation per cycle (R:W = 50:50) at the minimum operation voltage (*VDD_{min}*) at room temperature (RT).

4. CONCLUSION

As described in this paper, we proposed a read bitline amplitude limiter (RBAL) and discharge acceleration (DA) scheme. The RBAL reduces the active energy dissipation 13%–27% at 0.5 V. The RBAL increases the read delay. However, the DA scheme improves the delay overhead without power penalty. Although the delay overhead with RBAL and DA is 32% at worst case (FS, -40°C), the delay is decreased by 2%–5% at the SS corner. Circuits were implemented to 256-Kb SRAM macros by 40-nm process. The energy dissipation in a cycle is less than 10 pJ / access at 0.5–0.7 V.

5. ACKNOWLEDGMENTS

This work was carried out as a part of the Extremely Low Power (ELP) project supported by the Ministry of Economy, Trade and Industry (METI) and the New Energy and Industrial Technology Development Organization (NEDO).

6. REFERENCES

[1] A. J. Bhavnagarwala, X. Tang, and D. Meindl, "The Impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability," *Journal of Solid-State Circuits,* Vol. 36, No. 4, 658–665, 2001.

[2] R. Heald and P. Wang, "Variability in Sub-100nm SRAM Designs," Proc. of *International Conference on Computer Aided Design,* pp. 347–352, 2004.

[3] Y. Morita, K. Nose, K. Noguchi, S. Takami, K. Goto, Y. Aimoto, A. Kimura and M. Mizuno, "Small-defect detection in sub-100nm SRAM cells using a WL-pulse timing-margin measurement scheme," Digest of Technical Papers 2010 *Symposium on VLSI Circuits,* pp. 37-38, Jun. 2010.

[4] H. Pilo, J. Barwin, G. Braceras, C. Browning, S. Burns, J. Gabric, S. Lamphier, and M. Miller, "An SRAM Design in 65 nm and 45 nm Technology Nodes Featuring Read and Write-Assist Circuits to Expand Operating Voltage," IEEE *Symposium on VLSI Circuits,* pp. 15–16, June 2006.

[5] K. Takeda, T. Saito, S. Asayama, Y. Aimoto, H. Kobatake, S. Ito, T. Takahashi, K. Takeuchi, M. Nomura and Y. Hayashi, "Multi-step Word-line Control Technology in Hierarchical Cell Architecture for Scaled-down High-density SRAMs," Digest of Technical Papers 2010 *Symposium on VLSI Circuits,* pp. 101-102, Jun. 2010.

[6] Y. Fujimura, O. Hirabayashi, T. Sasaki, A. Suzuki, A. Kawasumi, Y. Takeyama, K. Kushida, G. Fukano, A. Katayama, Y. Niki, T. Yabe, "A Configurable SRAM with Constant-Negative-Level Write Buffer for Low-Voltage Operation with 0.149μm2 Cell in 32nm High-κ Metal-Gate CMOS," IEEE *International Solid-State Circuits Conference,* pp. 348-349, 2010.

[7] H. Nho, P. Kolar, F. Hamzaoglu, Y. Wang, E. Karl, Y. G. Ng, U. Bhattacharya and K. Zhang, "A 32nm High-κ Metal Gate SRAM with Adaptive Dynamic Stability Enhancement for Low-Voltage Operation," IEEE *International Solid-State Circuits Conference,* pp. 346-347, 2010.

[8] Y. Morita, H. Fujiwara, H. Noguchi, Y. Iguchi, K. Nii, H. Kawaguchi, and M. Yoshimoto, "An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design under DVS Environment," IEEE Symposium on *VLSI Circuits Digest of Technical Papers,* pp. 256–257, Jun. 2007.

[9] Y. Pu, X. Zhang, J. Huang, A. Muramatsu, M. Nomura, K. Hirairi, H. Takata, T. Sakurabayashi, S. Miyano, M. Takamiya, and T. Sakurai, "Misleading Energy and Performance Claims in Sub/Near Threshold Digital Systems," Proc. of *International Conference on Computer Aided Design,* pp. 625–631, 2010.

[10] A. Kawasumi, T. Suzuki, S. Moriwaki and S. Miyano, "Energy Efficiency Degradation Caused by Random Variation in Low-Voltage SRAM and 26% Improvement by Bitline Amplitude Limiting (BAL) Scheme," IEEE *Asian Solid-State Circuits Conference,* pp. 165-168, Sep. 2011.

[11] S. Yoshimoto, M. Terada, S. Okumura, T. Suzuki, S. Miyano, H. Kawaguchi and M. Yoshimoto, "A 40-nm 0.5-V 20.1-uW/MHz 8T SRAM with Low-Energy Disturb Mitigation Scheme," Digest of Technical Papers 2011 *Symposium on VLSI Circuits,* pp. 72-73, Jun. 2011.