56.67 fJ/bit single-ended disturb-free 5T loadless 4 kb SRAM using 90 nm CMOS technology

Chua-Chin Wang, Deng-Shain Wang & Sih-Yu Chen

Springer

Springer

CrossMark

56.67 fJ/bit single-ended disturb-free 5T loadless 4 kb SRAM using 90 nm CMOS technology

Chua-Chin Wang[1] · **Deng-Shain Wang**[1] · **Sih-Yu Chen**[1]

Abstract

A novel single-ended SRAM is proposed in this study, where the built-in self-refleshing data retention path has been utilized to reduce the SRAM cell area. In order to reduce the power-delay product, an analytical solution to derive the optimal number of the 5T cells on the BLB is reported in this paper. The proposed SRAM is implement by TSMC 90 nm CMOS technology. According to the measurement results, the energy dissipation per write/read operation is found to be 0.479/0.091 fJ provided that the SRAM cells is supplied a 0.6 V VDD supply.

1 Introduction

SRAM is often used as caches of processors for high speed computation such that it unavoidably consumes a significant portion of the entire power dissipation. To reduce the leakage power, a 4T loadless SRAM cell has been demonstrated in 2004 [1]. Two low-threshold NMOS are used as access transistors and two high-threshold PMOS are used as a latch-like storage so that the access time and the data retention can be enhanced. However, the degradation of the static noise margin (SNM) issue has reported by [2] since the SRAM cell lack of the bitline isolation mechanism. To improve the SNM, employ the variable bulk bias scheme reported in [3], however, the latch-up problem might be introduce as well. Charge recycling over bitlines was considered a method to reduce power dissipation [4]. The sacrifice of the speed, however, is not quite acceptable in certain cache applications. Notably, if the performance at low power supply voltage of the memory is able to meet the system demand, voltage scaling is considered the most effective alternative to reduce the power

dissipation therewith [5]. However, the price to pay of such a voltage scaling method is the long access time.

Although the loadless cell design was proved to be a very PDP-effective (power-delay product) solution for SRAM designs, its intrinsic "hidden self-recharging path" will not be reliable provided that nano-scale CMOS technology with supply voltage scaling is used. That is, the stored data bit (state) is likely to be damaged due to the reduction of supply voltage, the coupled noise, and smaller SNM [6, 7]. It is particularly serious when the access (R/W) operation is proceeding, where it will run into a disturb problem [8]. The asymmetrical W/R-assist approach proposed in [8], has drawn a great attention, since these works highlight the necessity of disturb-free design at the low voltage supply scenarios. Particularly, the disturb-free is highly demanded when the SRAM is required to carry out R/W in the subthreshold region using a nano-scale CMOS process. However, the area and power overheads caused by those auxiliary circuits and transistors are very significant. It is unacceptable in low power or low cost applications.

Even if the loadless SRAM cell with a write-assist loop to isolate the WLC was proposed to resolve the R/W access disturb issue such that the disturbance free is feasible [9], the performance has never been physically proven on silicon. Besides, the size and the number of the bitline drivers are not resolved. Thus, in this investigation, we physically implement the proposed 5T SRAM cell including a pair of

✉ Chua-Chin Wang
ccwang@ee.nsysu.edu.tw

1 Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan

Springer

high-V_{th} PMOS transistors to serve as the latch-like storage. Moreover, a shared read inverter is used to reject the potential bitline voltage variation so as to guarantee the disturb-free feature. We also derive the optimal solution for the number of cells on the same bitline to share the read inverter to achieve the best PDP performance analytically.

2 Distrub-free SRAM with loadless cells

2.1 4 kb SRAM design

Figure 1 shows the architecture of the entire 4 kb SRAM design, including the memory array, row decoder, column decoder, control circuit, Y-selector, and BIST (build-in-self test) circuit.

Two of the most widely used decoder designs are pseudo-NOR-gate-based and AND-gate-based. Since the former has a DC path consuming PC power, we employ the latter as the basic decoder design scheme to achieve low power dissipation.

2.2 Disturb-free loadless 5T SRAM cell

Although the prior loadless SRAM has been proven that the operating frequency is higher and the cell area is smaller than normal SRAM cells, the price to pay is the reduction of the noise margins, particularly SNM [6]. The reason, obviously, is caused by the bitline disturbance during the R/W operation. Such a disadvantage will become even worse if the SRAM is implemented using nano-scale technologies, where the leakage current, including the subthreshold current, tends to dominate the power dissipation and even compromise the stored data bits. We, then, propose a novel loadless SRAM cell in this study to resolve the above predicament. Referring to Fig. 2,
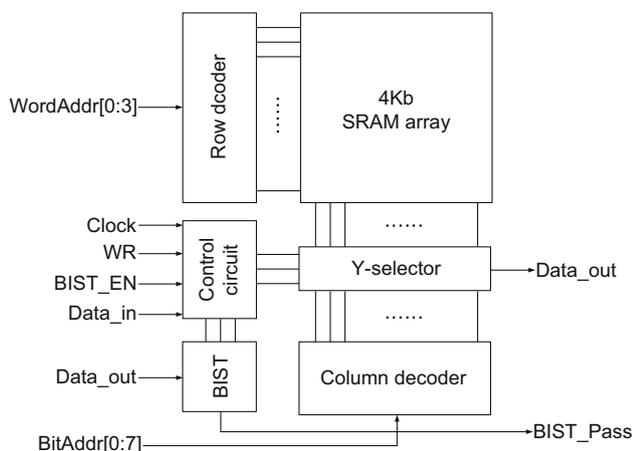
a single-ended disturbance-free load-free 5T SRAM cell is revealed.

Referring to Fig. 2, two high-V_{th} PMOS transistors, M201 and M202, consists of a latch-like storage, whereas low-Vth NMOS transistors, M203 and M204, are used as access switches. A high-V_{th} MOS has a lower turn-off leakage current than the normal MOS does, and vise versa, such that not only is data storage ensured stable, leakage currents from M203 and M204 is also reduced in this configuration. Besides, to reject the potential disturbance from the bitlines, we propose to insert one WL-controlled transistor (namely, WLC), namely M205, between BLB and the cell. The R/W access of the cell is as follows.

- *Read access* Figure 3 shows the read timing diagram of the proposed SRAM cell. As soon as the address lines are valid, the input signal Pre-discharge is pulled high to discharge the BLb to ensure the charge on BLb will not interfere the data during the read access. After discharging BLb, WL is asserted to turn on M205. Meanwhile, WE/RD is high and $\overline{WE/RD}$ is low. The state at Qb will be coupled to BLB via M204 and M205. Notably, a shared inverter is inserted between BLB and BL to reject the noise from BL. Besides, the shutoff M203 will ensure Q is free from the disturbance of BLB.

- *Write access* Figure 4 shows the write timing diagram of the proposed SRAM cell. When the address lines are valid, Pre-discharge will be logic 1 to discharge BLB. Then, WL is asserted to turn on M205. Meanwhile, according to Data_in, M204 (when Data_in is 1) or M203 (when Data_in is 0) will be turned on by WE/RD or $\overline{WE/RD}$ which has been pulled up to high. The state of Qb (when Data_in is 1) or Q (when Data_in is 0) will be coupled to BLB to either flip the state or stay the same. As stated earlier, the state of Q (or Qb) might become the complementary of Qb (or Q) via the latch-like PMOS pair. Most important of all, the shutoff M203 (or M204) prevents Q (or Qb) from the disturbance of BLB.

Therefore, the proposed loadless 5T SRAM cell is a single-ended disturb-free design. Not only can the R/W noise margins be enhanced, the area cost is also drastically reduced compared with the known disturb-free design approaches.

It is also well known that such a pseudo-latch still has the leakage problem which results in the loss of the stored data. The leakage can be neutralized by "hidden self-recharging path". Referring to an example in Fig. 5, assume that M203 is off. A total of 4 currents affect the voltage level of node Q when the data node Q is floating.

Fig. 1 Architectureof the proposed 4 kb SRAM

Fig. 2 Proposed disturb-free loadless SRAM cell (size unit: μm)
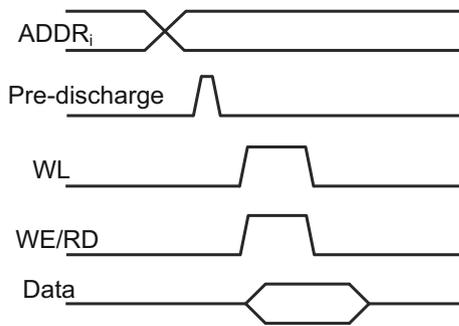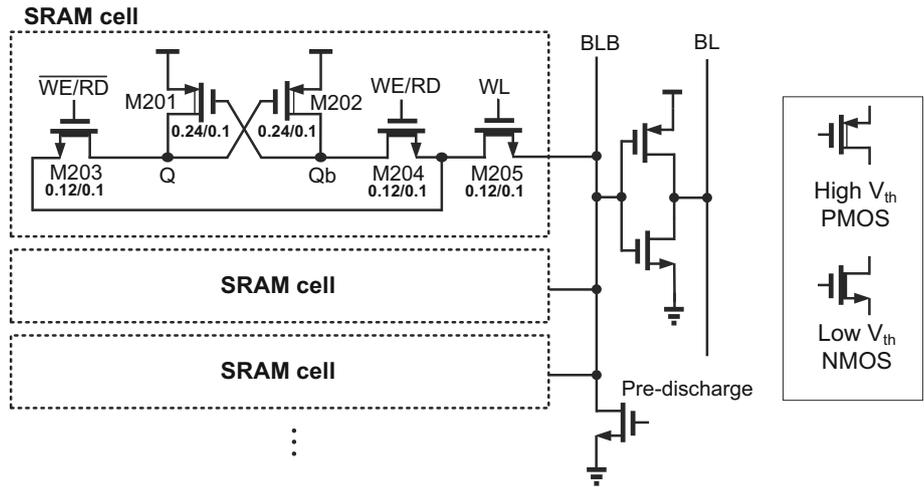
Fig. 3 Read timing diagram of the proposed SRAM cell
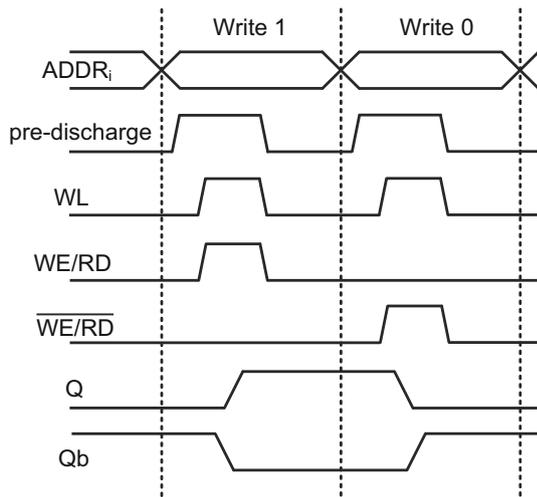
Fig. 4 Write timing diagram of the proposed SRAM cell

subthreshold current : I_{M201} and I_{M203}

reverse bias current : I_{D1} and I_{D2}

The requirement of the data retention for the possible weak "0" at node Q is $(I_{M201} + I_{D1}) < (I_{M203} + I_{D2})$. Notably, the
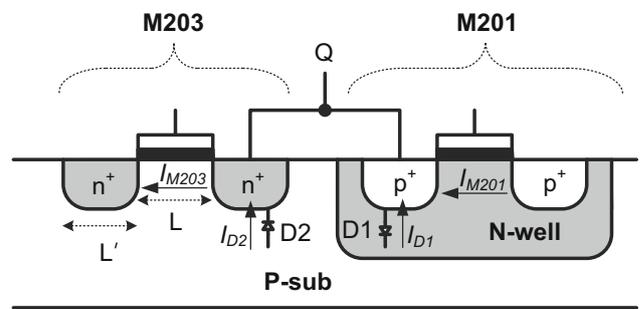
Fig. 5 Side view of the loadless SRAM cell

magnitude of the subthreshold currents are adjustable according to the following equations.

$$I_{sub} = \frac{W}{L} e^{\frac{V_{gs} - V_t}{nV_T}} \left(1 - e^{\frac{-V_{ds}}{V_T}}\right) \tag{1}$$

$$I_D = WL' \cdot I_S \left(e^{\frac{V}{V_T}} - 1\right) = I_{leakage} \tag{2}$$

where L and L' denote the length of the gate and the parasitic diode, respectively. Thus, the data retention problem can be resolved by tuning the W/L ratios to meet the requirement of $(I_{M201} + I_{D1}) < (I_{M203} + I_{D2})$. Apparently, the loadless SRAM cell attains the edge of area efficient based on the above analysis.

2.3 Optimal number of cells to share bitline inverter

Referring to Fig. 2 again, the inverter coupled between the bitlines (BL, BLB) is an important issue to be resolved. If the number of cells is large and the inverter size is small, the access time will be very long. On the contrary, if the number of cells is small and the inverter is large, unnecessary power will be consumed. In other words, the number of the cells to share this inverter as well as the bitlines must be determined before any physical implementation.

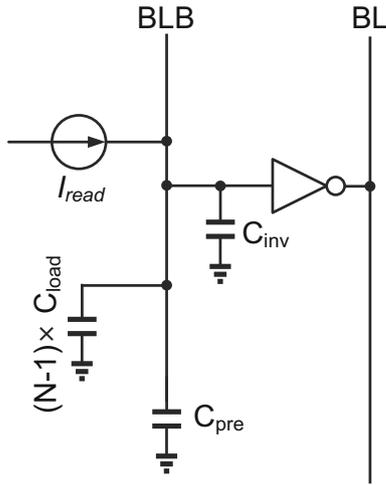**Fig. 6** Equivalent circuit model of the bitlines

Figure 6 is the equivalent schematic of Fig. 2, where M205 is deemed as the read current source, I_{read}, when the cell is selected to be accessed. By contrast, the other $(N-1)$ cells are seen as loads on BLB. Notably, the capacitance of the inverter and the discharge foot transistor is represented by C_{inv} and C_{pre}, respectively.

$$C_{load} = \epsilon'_{ox} \cdot W_{M205} \cdot L_{M205} \qquad (3)$$

where ϵ'_{ox} is the dielectric coefficient, W_{M205} and L_{M205} are the width and length of M205, respectively.

Thus, the overall capacitance on BLB is formulated as follows.

$$C_{total} = (N-1) \cdot C_{load} + C_{pre} + C_{inv} \qquad (4)$$

$$Power = f \cdot V_{DD}^2 \cdot C_{total} \qquad (5)$$

where N is the number of the cells coupled to the bitline, BLB, *Power* denotes the power dissipation on this bitline, f is the system read/write clock frequency, and V_{DD} is the supply voltage. Regarding the delay caused by the overall capacitance, it is formulated as follows.

$$\begin{aligned} Delay &= k_D \cdot \frac{C_{total}}{I_{read}} \\ I_{read} &= \frac{1}{2} \cdot \mu_0 \cdot \epsilon'_{ox} \cdot \frac{W_{M205}}{L_{M205}} \cdot (V_{GS} - V_{thn})^\alpha \end{aligned} \qquad (6)$$

where k_D is a constant, μ_0 is the mobility, α is the process parameter between 1.3 and 1.5 [1].

As addressed earlier, PDP is often used as an FOM to measure or evaluate the overall performance of circuit designs. The PDP of the proposed SRAM is as follows.

$$\begin{aligned} PDP &= Power \cdot Delay = f \cdot V_{DD}^2 \cdot C_{total} \cdot k_D \cdot \frac{C_{total}}{I_{read}} \\ &= \frac{f \cdot V_{DD}^2 \cdot [(N-1) \cdot C_{load} + Z]^2 \cdot k_D}{\frac{1}{2} \cdot \mu_0 \cdot \epsilon'_{ox} \cdot \frac{W_{M205}}{L_{M205}} \cdot (V_{GS} - V_{thn})^\alpha} \end{aligned}$$

where $Z = C_{pre} + C_{inv}$ for the sake of easy reading. The PDP cost function in the above can be re-arranged and simplified as the following equation.

$$\begin{aligned} PDP &= A \cdot W_{M205} + B + C \cdot \frac{1}{W_{M205}} \\ A &= \frac{2 \cdot f \cdot V_{DD}^2 \cdot (N-1)^2 \cdot \epsilon'_{ox} \cdot L_{M205}^3 \cdot k_D}{\mu_0 \cdot (V_{GS} - V_{thn})^\alpha} \\ B &= \frac{4 \cdot f \cdot V_{DD}^2 \cdot (N-1) \cdot L_{M205}^2 \cdot Z \cdot k_D}{\mu_0 \cdot (V_{GS} - V_{thn})^\alpha} \\ C &= \frac{2 \cdot f \cdot V_{DD}^2 \cdot L_{M205} \cdot Z \cdot k_D}{\mu_0 \cdot \epsilon'_{ox} \cdot (V_{GS} - V_{thn})^\alpha} \end{aligned} \qquad (7)$$

Referring to Eq. (7), the last term is negligible since it contains no $(N-1)$ and the width term is a reciprocal. Therefore, the PDP cost function is further simplified.

$$PDP \approx A \cdot W_{M205} + B \qquad (8)$$

The above conclusion predicts a linear relationship between PDP and the width of M205, namely the bitline driver of the proposed memory cell. We then carry out simulations provided that $N = 16, 32, 64$ to verify the derived PDP function. The outcome is shown in Fig. 7, where the prediction of such a linear relationship is proved. Besides, to reduce the PDP, the cell count N is 16 in the proposed work.

The above equation and simulations tell an interesting result. That is, the best PDP for the inverter-sharing bitlines will be governed by almost the minimal width of the
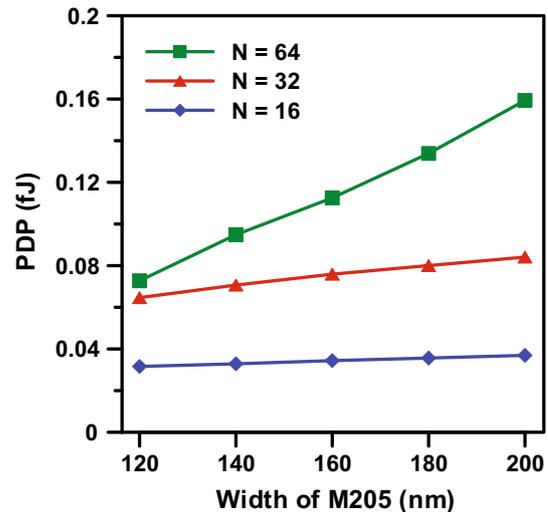

Fig. 7 PDP versus M205 transistor width (90 nm CMOS process)

bitline-driving transitor. Therefore, the entire disturb-free loadless SRAM is designed based on this viewpoint.

2.4 SRAM cell design and analysis

To ensure the functionality and performance of the proposed SRAM cell, all-PVT-corner simulations have been carried out. Figure 8(a, b) are the dynamic noise margin (DNM) and static noise margin (SNM), respectively, of the proposed SRAM cell. The DNM is around 0.3 V, which means that the state of the stored but won't be flipped even if the amplitude of the noise is high as 0.3 V given that the V_{DD} is 0.6 V. Regarding the SNM, Fig. 8(b) is not a traditional butterfly diagram due to it is a single-ended loadless cell. However, the SNM can be derived from the voltage difference that status of Q is not affected by the disturbance of Qb, where the minimal SNM, namely the worst case SNM, is 356.12 mV at FF and 100 °C corner.

2.5 BIST (built-in self-test)

The BIST circuit in the proposed is based on the March C-algorithm [10]. The algorithm is shown as follow :

$$\{\Updownarrow (w0); \Uparrow (r0, w1); \Uparrow (r1, w0); \Downarrow (r0, w1); \Downarrow (r1, w0); \Updownarrow (r0)\}$$

where $\Uparrow$ represents up count, $\Downarrow$ represents down count, $\Updownarrow$ represents up or down count, r means read, and w means write. March C-algorithm can detect the stuck-at fault (SAF), transition fault (TF), address-decoder fault (AF), and coupling fault (CF). If there is no error during the self-testing mode, the output pin, BIST_Pass, is set to logic 1.

3 Implementation and measurement

3.1 All-PVT simulations of disturb-free loadless 5T cell

The proposed SRAM is implemented by TSMC 90 nm CMOS mixed signal general purpose standard process.

Fig. 8 (a) Dynamic noise margin (DNM); (b) static noise margin (SNM) (at $W_{M205} = 0.12$ μm)
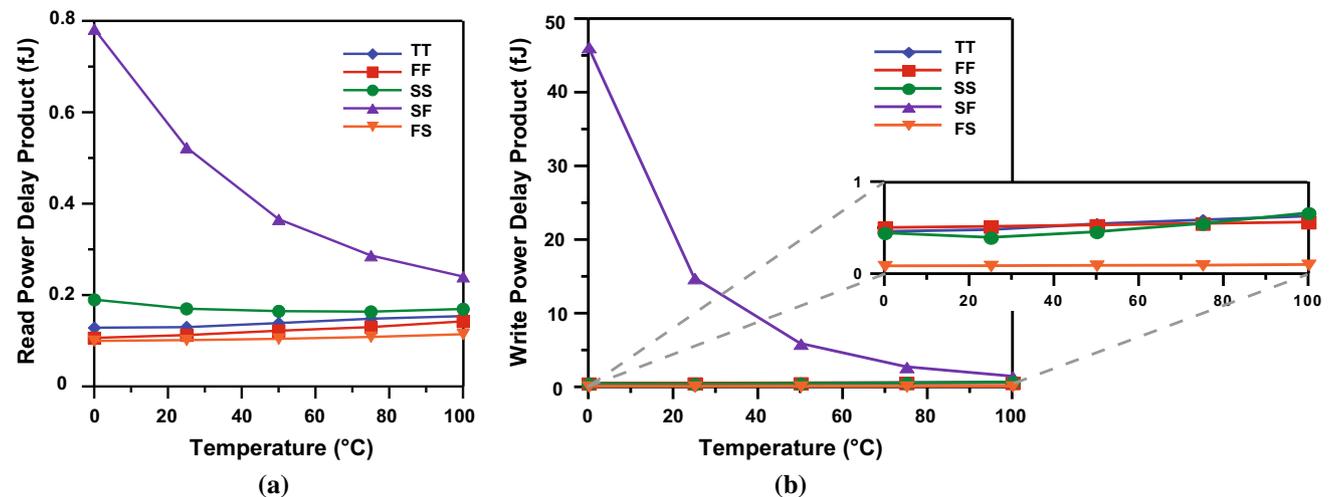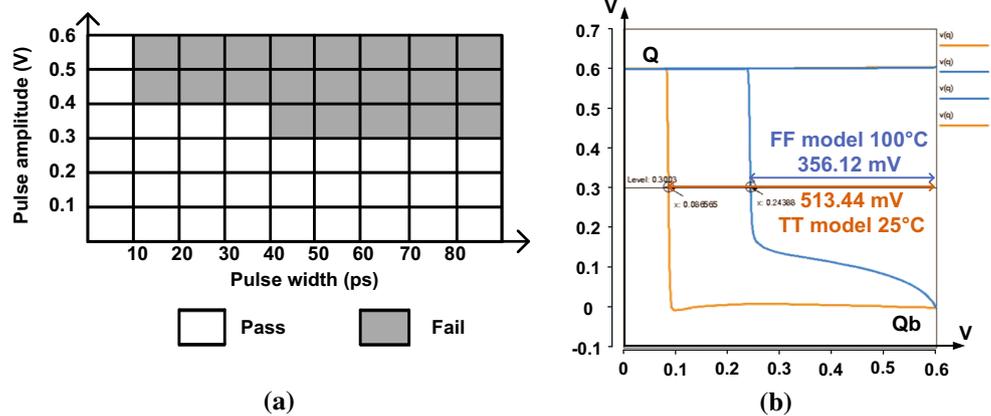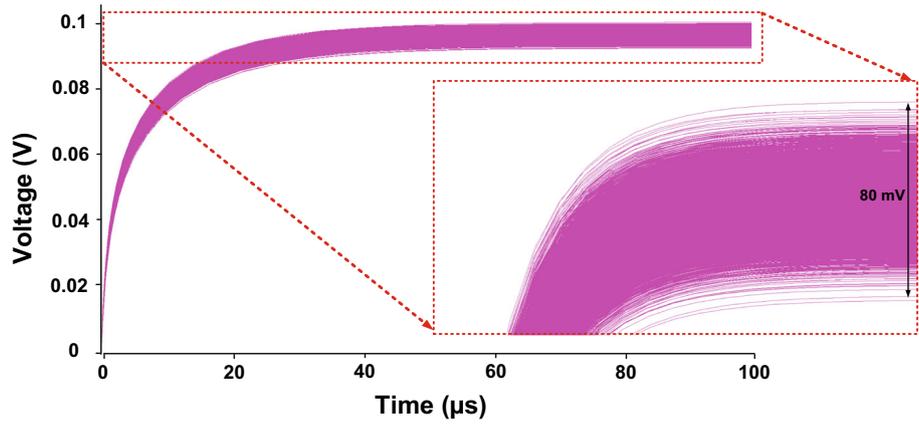
Fig. 9 PDP at different temperature and process corners. (a) Reading; (b) writing

Fig. 10 Monte–Carlo simulations of the proposed SRAM cell (10,000 times)

Fig. 11 a Layout; **b** die photo of the 4 kb SRAM

(a)

(b)
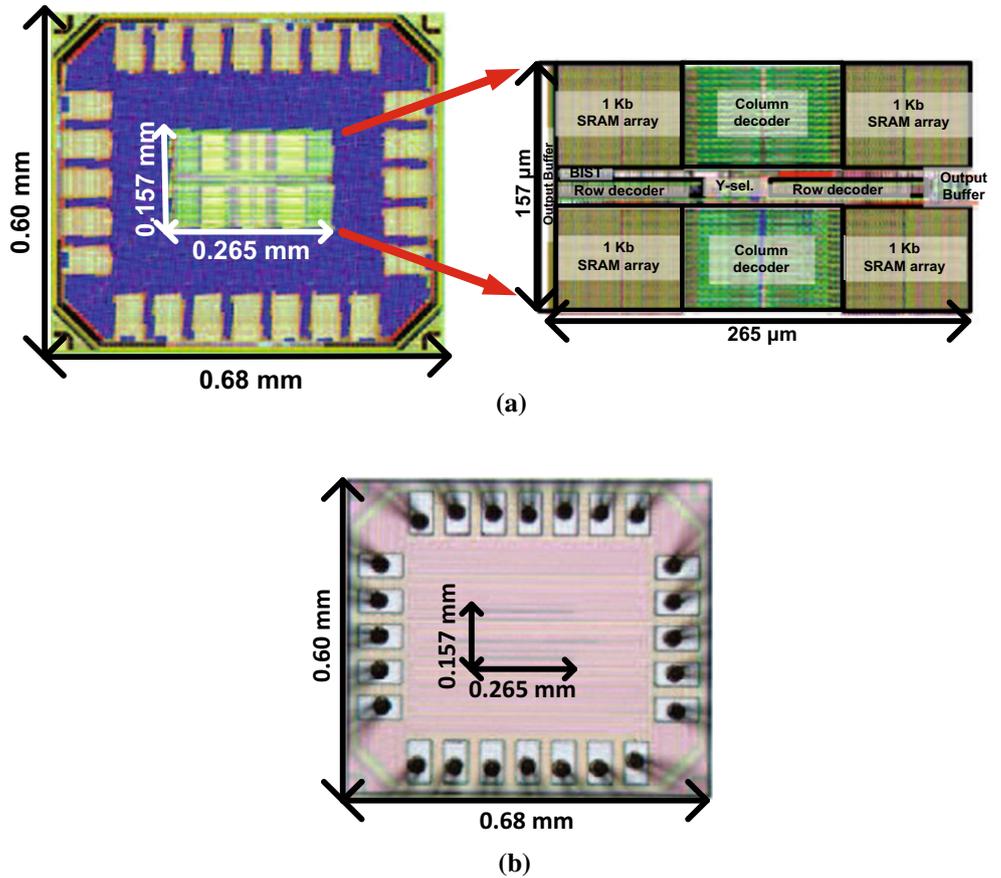
Referring to Fig. 9(a, b), which are the PDP distributions of reading and writing the cell at different temperature and process corners, respectively, the worst deviations are both found at SF corner (slow NMOS, fast PMOS). This is exactly expected since we employ NMOS as the bitline drivers and PMOS as the storage elements.

To analyze the effect of leakage, Fig. 10 shows 10,000 times Monte–Carlo simulations of the node Q of the SRAM cell which has been written 0. We assume that the voltage of node Q exceeding 0.1 V as a failure, because it

may cause long read delay and even wrong output. There is no failure in these 10,000 times of Monte–Carlo simulations.

3.2 Chip measurement

Figure 11 is the layout and die photo of the proposed 4 kb SRAM using the proposed 5T cells, respectively. The core area is 0.157×0.265 mm^2, while the overall area is 0.60×0.68 mm^2. The reason why the core circuit can not

Fig. 12 Random R/W access
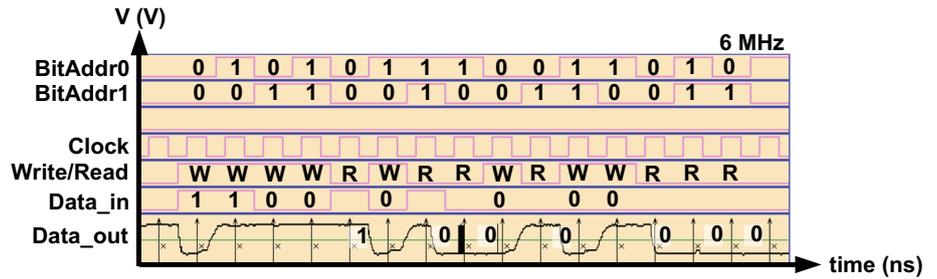

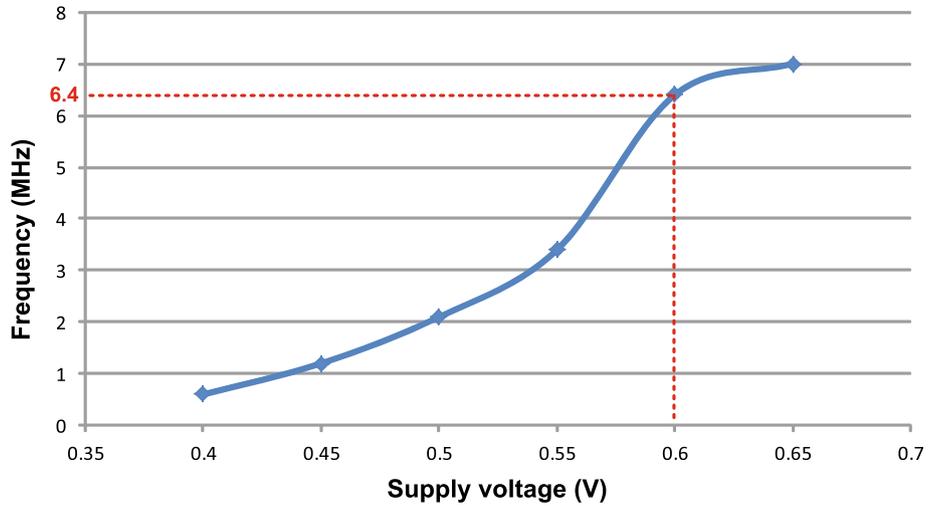**Fig. 13** Supply voltage versus the maximum operating frequency

Table 1 Standby and active power at different temperatures and supply voltage

	0 °C		25 °C		50 °C		75 °C		100 °C	
	Standby	Active	Standby	Active	Standby	Active	Standby	Active	Standby	Active
0.6 V	22.68	22.77	22.98	23.32	24.64	25.49	25.30	28.16	26.74	28.83
0.7 V	49.41	54.42	51.73	57.65	51.82	57.89	54.85	59.10	60.06	63.50
0.8 V	94.52	106.08	94.98	107.04	96.18	107.54	101.14	108.90	102.53	109.86
0.9 V	152.75	165.15	153.63	169.20	154.71	171.43	157.75	173.85	161.31	175.82
1 V	224.70	238.98	226.00	240.23	231.80	246.30	234.03	248.20	231.65	244.75

unit : μW

be shown in the diephoto is due to the requirement of metal density asked by the foundry such that the entire chip is almost covered with metal strips.

However, the most critical measurement of the proposed SRAM is the read and write function and performance. Figure 12 illustrates the random read and write access results, where the highest clock rate is 6.0 MHz. Figure 13 shows the relationship between the minimum supply voltage and the maximum operating frequency, where the operating frequency of the proposed SRAM is 6.4 MHz given 600 mV supply voltage. Table 1 tabulates the standby and active power when the proposed SRAM operates in different temperatures and supply voltages. At the room temperature, the standby power is 22.98 μW

provided that the supply voltage is 600 mV. The comparison with several prior works using the same 90 nm and better 65 nm CMOS process is summarized in Table 2. Our SRAM attains the lowest energy per bit compared with the prior works.

Notably, the proposed 4 kb SRAM using disturb-free loadless 5T attains the best energy/bit (word length = 1), the best normalized area/capacity, and the second best speed, compared with the recent SRAM designs fabricated by 90 and 65 nm process. We also like to highlight the fact that several function blocks in the 4 kb SRAM, e.g., Row decoder, Y-selector, and Column decoder in Fig. 1, are not deliberately realized with very lower power design techniques yet. That is, they are realized based on typical logic

Table 2 Comparison with prior works

	[11]	[12]	[13]	[14]	[15]	This work
Year	2014	2014	2015	2015	2016	2017
Process	40 nm CMOS	40 nm CMOS	22 nm FinFET	40 nm CMOS	65 nm CMOS	90 nm CMOS
Cell type	12T	8T	9T	5T	6T	5T
Supply voltage	0.35 V	0.65 V	0.3 V	0.6 V	1.2 V	0.6 V
Capacity	4kb	512 kb	32 kb	5 kb	1 kb	4 kb
Speed (MHz)	11.5	200	N/A	54	100	6
Energy/bit (fJ)	120	208	N/A	188.22	92.3	56.67
Standy power (μW)	N/A	107	2.1	N/A	N/A	22.98
Core size (μm²)	132 × 132	N/A	N/A	137 × 182	209 × 61	157 × 265
Normalized Area/Capa.	26.59	31.72	59.1	28.3	29.47	12.54

$$\text{Normalized Area/Capa.} = \frac{\text{core area}}{\text{capacity}} \times \frac{10000}{\text{process}^2(\text{nm}^2)}$$

design methodology. Otherwise, the overall PDP as well as the energy/bit will be reduced even more significantly.

4 Conclusion

This paper presents a 4 kb SRAM on silicon using the proposed low-power single-ended loadless 5T SRAM cell with disturb-free access. In particular, the potential interference from bitlines is de-coupled during the access by inserting a WL-controlled NMOS between the cell and BLB. Most important of all, an analytical solution to derive the optimal number of the 5T cells on the BLB is reported such that the PDP performance, or energy/bit, of the proposed design is expected to be the best.

References

1. Wang, C.-C., Tseng, Y.-L., Leo, H.-Y., & Hu, R. (2004). A 4-Kb 500-MHz 4-T CMOS SRAM using low-V_{THN} bitline drivers and high-V_{THP} latches. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(9), 901–909.
2. Wang, C.-C., Lee, C.-L., & Lin, W.-J. (2007). A 4-Kb low power SRAM design with negative word-line scheme. *IEEE Transactions on Circuits & Systems I : Regular Papers*, 54(5), 1069–1076.
3. Wang, C.-C., Sung, G.-N., Lee, C.-L., Chen, T.-H., Lin, W.-J., & Hu, R. (2008). 1.7-ns access time SRAM using wordline-controlled transistors with variable bulk bias. *Journal of Circuits, Systems, and Computers (JCSC)*, 17(5), 943–956.
4. Yang, B.-D. (2010). A low-power SRAM using bit-line charge-recycling for read and write operations. *IEEE Journal of Solid-State Circuits*, 43(2), 2173–2183.
5. Morifuji, E., Yoshida, T., Kanda, M., Matsuda, S., Yamada, S., & Matsuoka, F. (2006). Supply and threshold-voltage trends for scaled logic and SRAM MOSFETs. *IEEE Transactions on Electron Devices*, 53(6), 1427–1432.
6. Seevinck, E., List, F. J., & Lohstroh, J. (1987). Static-noise margin analysis of MOS SRAM cells. *IEEE Journal of Solid-State Circuits*, SC–22(5), 748–754.
7. Makino, H., Nakata, S., Suzuki, H., Mutoh, S., Miyama, M., Yoshimura, T., et al. (2012). Utilising the normal distribution of the write noise margin to easily predict the SRAM write yield. *IET Circuits, Devices & Systems*, 6(4), 260–270.
8. Vaalaee, A., & Al-Khalili, A. J. (2012). High-performance low-power sensing scheme for nanoscale SRAMs. *IET Computers & Digital Techniques*, 6(6), 406–413.
9. Chen, S.-Y., & Wang, C.-C. (May 2012). Single-ended disturb-free 5T loadless SRAM cell using 90 nm CMOS process. In *IEEE international conference on IC design and technology* (pp. 1–4).
10. Al-Harbi, S., & Gupta, S. (April 2001). An efficient methodology for generating optimal and uniform march tests. In *Proceedings of IEEE VLSI test symposium* (pp. 231–237).
11. Chiu, Y.-W., Hu, Y.-H., & Tu, M.-H. (2014). 40 nm bit-interleaving 12T subthreshold SRAM with data-aware write-assist. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61(9), 2578–2585.
12. Lien, N.-C., Chu, L.-W., Chen, C.-H., & Yang, H.-I. (2014). A 40 nm 512 kb cross-point 8 T pipeline SRAM with binary word-line boosting control, ripple bit-line and adaptive data-aware write-assist. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61(12), 3416–3425.
13. Yang, Y., Juhyun, P., Song, S.-C., Wang, J., Geoffrey, Y., & Jung, S.-O. (2015). Single-ended 9T SRAM cell for near-threshold voltage operation with enhanced read performance in 22-nm FinFET technology. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 23(11), 2748–2752.
14. Wang, C.-C., Wang, D.-S., & Liao, C.-H. (2015). A leakage compensation design for low supply voltage SRAM. *IEEE*

Transactions on Very Large Scale Integration (VLSI) Systems, *24*(5), 1761–1769.

15. Yang, Y., Juhyun, P., Song, S.-C., Wang, J., Geoffrey, Y., & Jung, S.-O. (May 2016) A 17.5-fJ/bit energy-efficient analog SRAM for mixed-signal processing. In *IEEE international symposium on circuits and systems* (pp. 22–25).

Chua-Chin Wang received Ph.D. degree in electrical engineering from SUNY (State University of New York) at Stony Brook, USA, in 1992. Dr. Wang's research interests include memory and logic circuit design, communication circuit design, and interfacing I/O circuits. He was elevated to be Distinguished Professor of National Sun Yat-Sen University in 2010. He became IET Fellow in 2012. Dr. Wang was General Chair of 2007 VLSI/CAD Symposium. He was General Co-Chair of 2010 IEEE Inter. Symp. on Next-generation Electronics (2010 ISNE). He was General Chair of 2011 IEEE Inter. Conf. on IC Design and Technology (2011 ICICDT), and General Chair of 2012 IEEE Asia-Pacific Conference on Circuits & Systems (2012 APCCAS).

Deng-Shain Wang was born in Taiwan in 1988. He received the B.S. and M.S. degrees in electronic engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan, in 2011 and 2013, where he is currently pursuing the Ph.D degree in electrical engineering. His recent research interest focuses on analog design.

Sih-Yu Chen was born Taiwan in 1989. She received M.S. degree in electrical engineering from National Sun Yat-Sen University, Taiwan in 2013. Her recent research interest is focused on analog design.