A 0.506-pJ 16-kb 8T SRAM With Vertical Read Wordlines and Selective Dual Split Power Lines

Lu Lu, *Student Member, IEEE*, Taegeun Yoo, *Member, IEEE*, Van Loi Le, *Student Member, IEEE*, and Tony Tae-Hyoung Kim, *Senior Member, IEEE*

***Abstract*— This article presents an 8T static random access memory (SRAM) macro with vertical read wordline (RWL) and selective dual split power (SDSP) lines techniques. The proposed vertical RWL reduces dynamic energy consumption during read operation by charging and discharging only selected read bitlines (RBLs). The data-aware SDSP technique combined with vertical write bitlines enhances both the write margin (WM) and the static noise margin (SNM). A 16-kb SRAM test chip fabricated in 65-nm CMOS technology demonstrates the minimum energy consumption of 0.506 pJ at 0.4 V and the minimum operating voltage of 0.26 V.**

***Index Terms*— Data aware, low power, selective dual split power (SDSP) lines, static random access memory (SRAM), vertical read wordline (RWL).**

I. Introduction

With the recent development of portable devices and wearable apparatus requiring long battery lifetime, ultralow power consumption has been increasingly becoming important. In their performance, power, and energy, static random access memories (SRAMs) play a vital role [1]. As a crucial approach to meet ultralow power and energy, supply voltage scaling has been commonly employed. However, SRAMs with smaller-sized transistors restrict the minimum operating voltage due to various key metrics such as stability, sensing margin, write margin (WM), and so on. In conventional 6T SRAMs, the minimum operating voltage is limited by the conflicting requirements from write ability and read stability [2], [3]. Therefore, SRAM cells with decoupled read ports are widely employed for supply voltage scaling since they can eliminate read disturbing current and allow more flexibility in the cell design [4]. In addition, SRAMs operating in near- or sub-threshold region have significantly degraded I_{ON}-to-I_{OFF} ratios and exponentially increased variations, which deteriorates many SRAM design parameters. Many design techniques have been reported to partially address these SRAM design issues [5]–[15].

For a read operation, several techniques could be used to enhance the I_{ON}-to-I_{OFF} ratio between active discharge current and leakage current including: 1) increasing the discharge current such as boosted (BST) read wordline (RWL) on select cells or using negative cell virtual ground voltage [5]; both incur the power and area overhead; 2) providing a pull-up current against the bitline leakage to improve the sensing margin and sensing window (such as using an extra transistor to equalize the leakage [6]); however, the bitline leakage is equalized to maximum leakage, which increases power consumption; 3) adopting hierarchical bitline to mitigate cell disturbance by reducing the amount of loading [7]; and 4) generating the self-adjusted reference voltage for sensing amplifiers to enhance the sensitivity to small bitline swing [8]. At higher supply voltage, the dynamic power contributes the most consumption in the read operation. In the conventional SRAMs, the majority of the switching power is consumed to precharge the unselected bitlines due to the horizontal RWLs [6]–[8]. The proposed vertical RWLs reduce the dynamic power significantly by eliminating the current flowing in unselected columns. At lower supply voltage, the leakage power becomes equally significant in the total power consumption. To address this issue, higher threshold voltage (HVT) devices are used in the noncritical paths to reduce the leakage current without performance degradation. Virtual ground is widely used to suppress bitline leakage current. In [9], raised virtual ground and floating bitlines with programmable nMOS sleep transistors are utilized to reduce the leakage current.

For a reliable write operation, write access transistors need to be stronger than pull-up pMOS transistors in SRAM cells. However, the conflicting requirements between static noise margin (SNM) and WM restrict supply voltage scaling [10]. Various techniques have been reported to ameliorate these issues. For example, wordline underdrive (WLUD) improves SNM while degrading WM [11], [12]. On the contrary, lowering cell supply voltage [13], [14] and wordline overdrive (WLOD) and raising VSS will enhance WM while leading to deteriorated SNM. Another technique is using negative bitline (NBL) [13], [14] for WM enhancement. However, this suffers from area penalty caused by negative voltage generation and power overhead. Therefore, most studies combine two or more techniques [15], [16] to achieve the target SNM and WM. In [16], split control of write access transistors is proposed

Manuscript received June 25, 2019; revised October 7, 2019; accepted November 5, 2019. Date of publication April 8, 2020; date of current version June 1, 2020. *(Corresponding author: Lu Lu.)*

L. Lu and V. L. Le are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: llu010@e.ntu.edu.sg).

T. Yoo is with VIRTUS, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

T. T.-H. Kim is with the Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2019.2956232

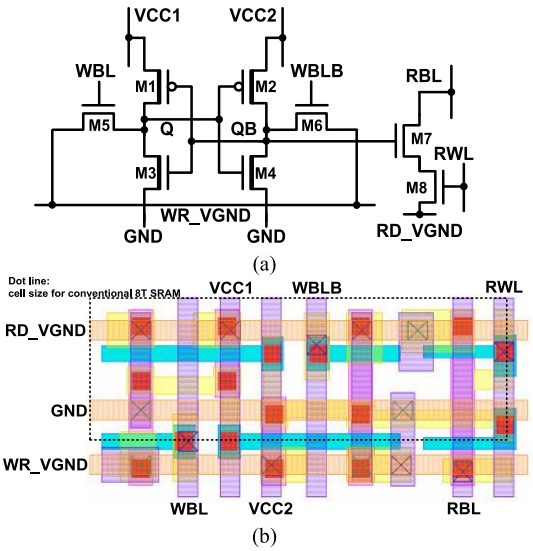


Fig. 1. (a) Proposed 8T SRAM cell with HVT devices in the latch and RVT devices in the access transistors and the read port. (b) Layout of the 8T SRAM cell based on the 65-nm logic design rules.

TABLE I
VERTICAL RWL READ OPERATION

	Selected	Unselected	Holding
RD_VGND	Low	High	High
RWL	High	Low	Low

while raising VSS and NBL to enhance the write ability. Meanwhile, the SNM for inactive cells is weakened. Lutkemeier *et al.* [17] proposed vertical power lines to improve the write ability also at the cost of stability degradation. The concept of split power line with body bias control was reported in [18]. It improves WM and SNM with significant overhead in the cell area. Overall, no simple technique can improve WM without SNM degradation and vice versa. Hence, it is highly necessary to develop techniques for improving WM without SNM degradation or enhancing SNM without WM degradation.

This article presents two techniques to reduce energy consumption and improve SNM and WM. The vertical RWL reduces the read dynamic energy. The selective dual split power (SDSP) with selective boosted cell supply improves both WM and SNM compared to the conventional 6T or 8T SRAMs [19]. The remainder of this article is organized as follows. Section II explains the key principles of the proposed design. Measurement results are presented in Section III. Finally, the conclusion is drawn in Section IV.

II. PROPOSED SRAM ENERGY CONSUMPTION IMPROVEMENT TECHNIQUES

A. Proposed SRAM Cell

Fig. 1(a) shows the schematic of the proposed 8T SRAM cell. The 8T SRAM cell consists of a cross-coupled latch (M1–M4), write access devices (M5 and M6), and a read port (M7 and M8). The RWL, the read bitline (RBL), and the write bitlines (WBL and WBLB) are controlled vertically, whereas the read virtual ground (RD_VGND) (Table I) and the write virtual ground (WR_VGND) are controlled horizontally. M7 is connected to Q bar (QB) instead of RWL to control the coupling capacitor between RWL and RBL. In a write operation, WBL, WBLB, VCC1, and VCC2 are biased separately depending upon the write data (Table II). WR_VGND of the selected cell is grounded, whereas those of the unselected cells are precharged to VDD and left floating. The cross-coupled latch is implemented with HVT devices for leakage reduction, whereas the write access paths and the read port employ regular threshold (RVT) devices for better performance. The novelty of the proposed 8T SRAM cell, when compared to the conventional 8T SRAM cell, includes the vertical RWL and the data-aware control of VCC1, VCC2, WBL, and WBLB. The detailed operation of the cell will be explained in Section II-B.

TABLE II
DATA-AWARE WRITE OPERATION IN THE SELECTED CELL

	Writing '0'	Writing '1'	Unselected	Holding
WBL	High	Low	Low	Low
WBLB	Low	High	Low	Low
WR_VGND	Low	Low	Floating	High
VCC1	High	Boosted	High	High
VCC2	Boosted	High	High	High

Fig. 1(b) shows the cell layout of the proposed 8T SRAM cell. The placement of the transistors is the same as the conventional 8T cell [20] and follows a similar layout except the metal routing. The power lines in the conventional cell are split into VCC1 and VCC2 running vertically. RD_VGND and WR_VGND are implemented with horizontal metal-3 lines, and RWL is realized in a vertical metal-2 line. The vertical metal congestions cost 0.9% additional power consumption. RWL and RBL are placed next to each other since the cell area is compact. The effective coupling capacitor is 14 fF. The cell layout occupies $2.23 \times 1.02\ \mu m^2$ when designed in the logic rule of the employed 65-nm CMOS technology, which is 1.45 times of the conventional 8T SRAM cell.

B. Energy Reduction With Vertical RWL

SRAM energy consists of dynamic (switching) and static energy (leakage). Several techniques are available for leakage reduction. Floating bitlines [9] could reduce the leakage current through access transistors by automatically generating bitline voltage providing minimum bitline leakage. Body biasing [21] is another method for leakage reduction. Since leakage current is an exponential function of the device threshold voltage, applying reverse body bias can reduce leakage current effectively. Another common idea is to adopt multiple V_{th} devices multiple threshold CMOS (MTCMOS) [22] so that the leakage in the noncritical paths can be reduced. Virtual ground (RD_VGND) [9] is also widely used to reduce the leakage flowing through read ports. This article also employs MTCMOS and RD_VGND to reduce the leakage in the hold mode.

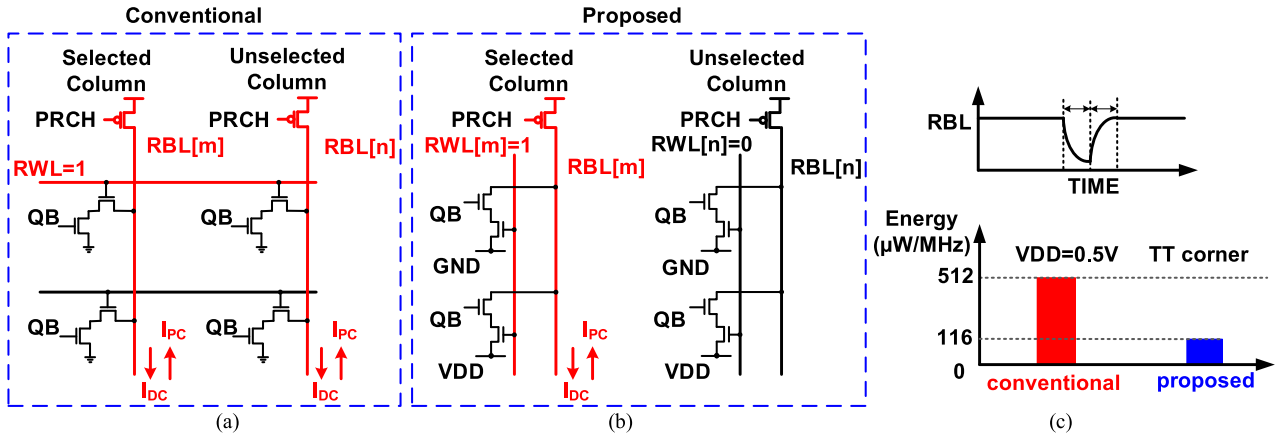


Fig. 2. (a) Conventional read operation in 8T SRAM. (b) Proposed read operation in 8T SRAM. (c) Energy consumption comparison of conventional one and proposed one. Note that the array size of 256 rows × 128 columns with the multiplexer ratio of 16-to-1 is assumed.

One most significant dynamic read energy in the conventional SRAMs comes from bitline discharging and precharging. Even though only a few columns are accessed for read operation, activating a horizontal RWL can allow all RBLs to be discharged in the worst case. And since a significant amount of variations require large guard bands in time, horizontal RWL tends to fully discharge RBLs to ground [23]. This article tackles this issue by utilizing a vertical RWL scheme. Fig. 2 compares the energy consumption of the proposed RWL scheme with that of the conventional counterpart. Note that the array size of 256 rows × 128 columns with the multiplexer ratio of 16-to-1 is assumed. In the conventional scheme [Fig. 2(a)], all RBLs are discharged in the worst-case scenario where all QBs in the selected row are "1." This is not the case in the proposed SRAM where only a selected RBL can be discharged to a lower level. This is possible since the vertical RWLs of the unselected columns are grounded [Fig. 2(b)]. This reduces the read energy significantly. In the proposed SRAM, the discharging current in the selected column is the pull-down cell current minus the pull-up leakage current. In the conventional SRAM, the discharging current consists of the pull-down cell read current and the pull-down leakage current. The dynamic discharging energy in the selected column of the proposed scheme is 10%–35% larger than that of the conventional one. Since RD_VGND is used, it increases leakage by 17%, power by 3%, and area by 4.8%. However, the proposed SRAM eliminates the dynamic energy from the unselected columns. As shown in Fig. 2(c), in the SRAM array, assuming the data pattern is 50% "1" and 50% "0," under room temperature and 0.5-V supply voltage, the energy consumption in the array without peripheral circuit for the conventional read port is five times that of the proposed one. Therefore, the overall energy of the proposed SRAM is smaller than that of the conventional 8T SRAM.

C. RBL Sensing With Vertical RWL

A reliable read operation depends on the sensing margin, which is defined as the difference between the RBL levels of data "0" and "1." The RBL level is determined by several parameters such as bitline leakage, bitline length, supply voltage,

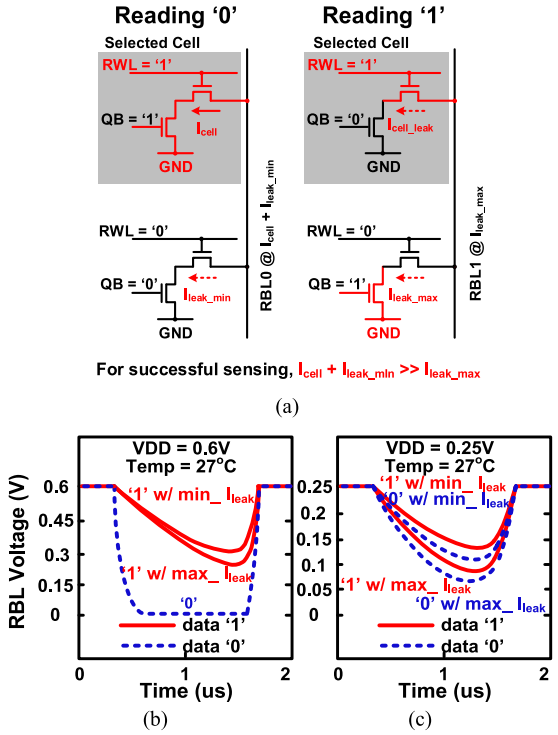


Fig. 3. (a) Principle of the conventional read operation. (b) Sensing margin at 0.6 V. (c) Sensing margin at VDD = 0.25 V. The number of cells per bitline is 256.

and so on. In the conventional SRAM [Fig. 3(a)], the worst case scenario for reading "0" is with minimum pull-down leakage current, which occurs when all QBs in the unselected cells are "0s." This produces the lowest RBL discharging speed. Likewise, the worst case scenario of reading "1" is with maximum leakage current, which occurs when all QBs in the unselected cells hold data "1." Maximum pull-down leakage discharging RBL for data "1." The necessary condition for successful reading is that the cell discharging current with the minimum leakage current ($I_{cell} + I_{leak\_min}$) needs be larger than the maximum leakage current (I_{leak_max}). Supply voltage scaling will make I_{cell} close to I_{leak} and degrades the

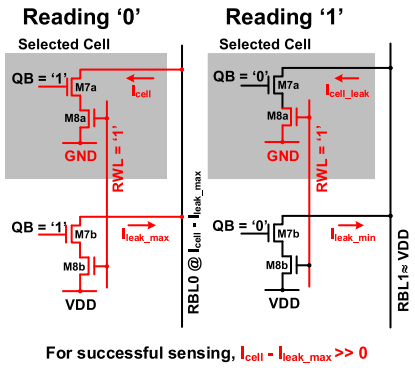


Fig. 4. Principle of the proposed read operation.

sensing margin. Fig. 3(b) and (c) shows the impact of supply scaling and bitline leakage on RBL sensing. At higher level supply voltage [Fig. 3(b)], the I_{ON}-to-I_{OFF} ratio is so high that the pull-down cell current is much larger than the bitline leakage. Therefore, the sensing margin is large enough for single-ended sense amplifiers to be adopted. However, when the supply voltage scales to the subthreshold region, bitline leakage becomes close to the pull-down cell current. This increases the impact of the bitline leakage on the sensing margin. As presented in Fig. 3(c), the RBL level for data "1" could be lower than that of data "0" because of the significant data-dependent bitline leakage.

Fig. 4 illustrates the worst case read conditions of the proposed vertical RWL technique. When the read data is "0" (QB = "1"), RD_VGND of the selected row is grounded, and RBL is discharged through M7a and M8a. The half-selected cells in the same column generate pull-up leakage and prevent RBL from being fully discharged to the ground. When all the half-selected cells store "0" (QB = "1"), the pull-up leakage will be maximized forming the highest RBL level for data "0." Similarly, if the read data is "1," the worst case RBL level is found when all the half-selected cells store "1" (QB = "0"). This will lead to the lowest RBL level for "1." Unlike the conventional SRAMs only with pull-down current paths, the proposed scheme has both pull-down and pull-up current paths in RBL. This forms static RBL levels for data "0" and "1," which widens the sensing timing window. As mentioned in Section II-A, parallel RWL and RBL create coupling capacitance, which affects RBL sensing during the read operation. In this article, the coupling capacitor between RWL and RBL boosts the RBL voltage and increases the sensing margin [24]. However, when M7 and M8 are in conventional positions, the coupling effect becomes too strong and slows down RBL sensing. To address this issue, the locations of M7 and M8 are swapped, which demonstrates a better sensing margin. Fig. 5(a) shows that the proposed RBL has better sensing margins or sensing speed compared to the swapped cell without metal line coupling and the vertical RWL without M7 and M8 swapping. The sensing margin decreases with increasing number of cells on RBLs. When the loading is 256 per RBL, the margin still meets the target sensing margin of >100 mV at 400 mV even after including process voltage temperature (PVT) variations. Fig. 5(b) compares the RBL waveforms of the proposed vertical RWL with that of the

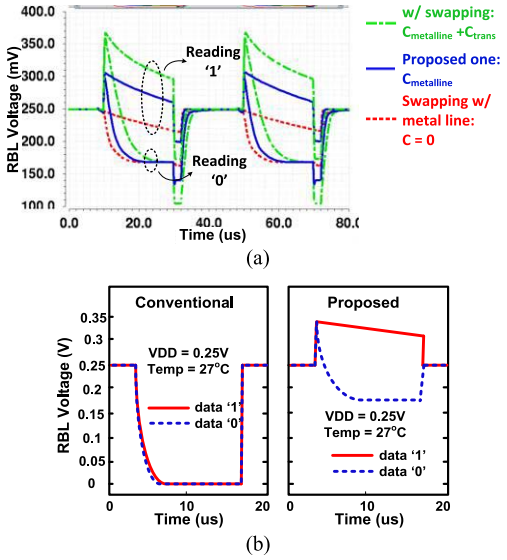


Fig. 5. (a) RBLs with different coupling capacitors under 0.25 V. (b) Impact of the proposed vertical RWL on the RBL swing when compared with the conventional RBL structure.

conventional scheme. It is obvious that the proposed scheme produces better sensing timing window and sensing margin compared to the conventional 8T SRAM illustrated in Fig. 3.

Fig. 6(a) shows the read time comparison between the proposed vertical RWL scheme and the conventional horizontal RWL scheme. The column data pattern is 50% "0" and 50% "1." Fig. 6(b) presents the normalized read time. The read speed is mainly decided by the discharging current on the selected cell. In this article, the sensing margin is large enough before the RBL becomes saturated to VDD-V_{th}. Therefore, the impact of the proposed vertical RWL on the read delay is insignificant. Overall, the proposed read time is slightly larger than that of the conventional one. At 0.4 V and 125 °C, the ratio reaches 1.56.

Fig. 6(c) shows the read energy consumption comparison in a subarray (16 columns) without a peripheral circuit. It can be seen that the energy of the conventional SRAM is more sensitive to the data pattern. Note that all the columns can consume energy in the conventional SRAM, whereas only the selected columns can dissipate energy in the proposed SRAM. Therefore, the data dependency of the proposed SRAM is much smaller than that of the conventional SRAM. With the data pattern of 50% "0" and 50% "1" and 1-V supply voltage, the energy consumption of the conventional horizontal RWL SRAM is five times as large as that of the proposed SRAM. When the supply voltage is 0.4 V, the proposed SRAM consumes one-third of the energy consumed by the conventional SRAM. Considering the peripheral logic, the energy consumption is 49.7% and 73% of the conventional SRAM for all data is "0" and 50% "0"/50% "1" data pattern, respectively. This demonstrates that the proposed vertical RWL scheme reduces energy significantly.

D. Sensing Amplifier With the Reference Signal

The RBL of the proposed SRAM cannot be fully discharged to GND due to the pull-up leakage current. This prevents

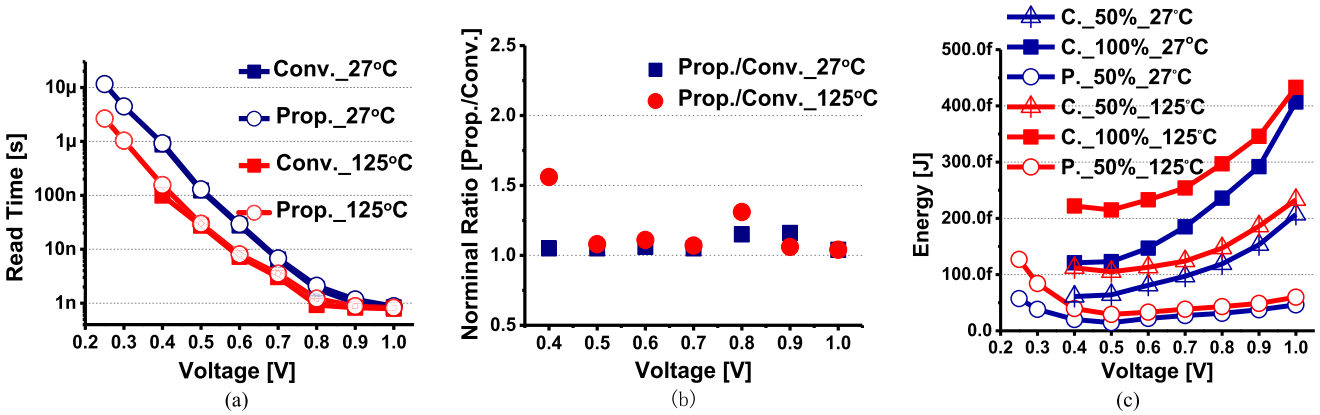


Fig. 6. (a) Read times of the proposed and conventional, rows per bank (RPB) = 256, CMUX = 16, in typical typical (TT) corner with varying temperature and supply voltage. (b) Normalized read time in varying voltage and temperature. (c) Read energy consumption of the proposed and conventional one, RPB = 256, CMUX = 16, in TT corner with varying temperature and supply voltage. Note that these results compare only SRAM arrays without peripheral circuits.

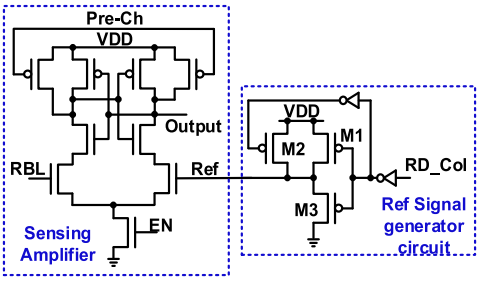


Fig. 7. Diagram of the sensing amplifier and reference signal generator.

simple inverters from being employed as sense amplifiers. Fig. 7 illustrates the sense amplifier employed in this study with reference to signal generation. The output of the sense amplifier is precharged to VDD during standby. During the read operation, the RBL voltage is compared with the reference signal (Ref) that is generated from a reference generator. When the read column signal (RD_Col) is "low," M2 is turned ON and maintains Ref "high." Once RD_Col rises up, M2 is open and the current flows through M1 and M3 as a voltage divider to generate Ref. The reference generator is realized with three pMOS devices whose sizes are determined through comprehensive simulation.

E. Read-Assists Comparison

In Table III, the proposed read port is compared to the conventional 8T SRAM with some related works. They all use the horizontal word line to control the selective transistor; in the worst scenario, large energy is consumed by discharging and precharging all the RBLs. Especially the design in [26] uses differential voltage sensing, and irrespective of the data pattern in the row, it discharges all RBLs. However, only the selected RBL is discharged in the proposed scheme. For the work in [24], Mos cap is used to boost the RBL, and lowering cell VDD is adopted to improve the sensing margin, both are power-hungry techniques. In the proposed idea, the coupling capacitor between the metal lines is used to boost the RBL, which is energy saving.

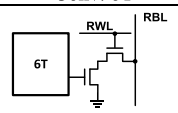
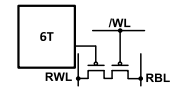
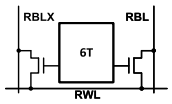
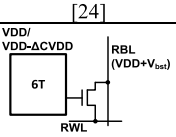
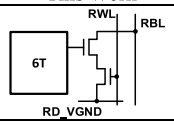
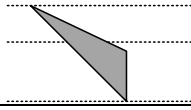
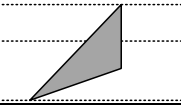
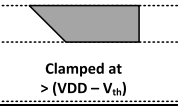
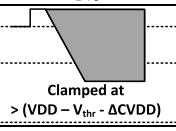
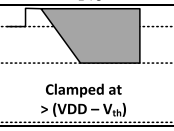
III. Proposed SRAM Ultralow Voltage Techniques

A. Write Operation With SDSP Lines

HVT devices are adopted at the cross-coupled latch of the proposed SRAM cell to reduce the leakage current. However, this degrades the write speed at ultralow voltage operation, which makes the write speed dominant and increases the overall energy. To ameliorate this issue, we adopt RVT devices on the pass-gate transistors. However, this will deteriorate the stability of the half-selected cells in a selected row. The split power line in [18] improves WM by sacrificing stability. Body bias control with a significant area overhead in the cell is employed to compensate for the stability degradation. In this article, we utilize data-aware VCC1/2 control to improve both write speed and WM. Furthermore, split-boosted power lines are proposed to improve the SNM of the half-selected cells. Therefore, the proposed SRAM can operate reliably at an ultralow supply voltage, consequently reducing the total write energy.

Fig. 8 illustrates the principle of the proposed SRAM write operation with SDSP, and Fig. 9 indicates the approach of VCC1/2 to improve the write ability and stability using written "0" as an example. In the selected row, WR_VGND[0] is connected to GND. If the write data is "0," WBL[0] is raised to VDD and Q_{00} is discharged to WR_VGND[0] [Fig. 8(a)]. Meanwhile, a charge pump circuit generates boosted voltage on VCC2[0]. As shown in Fig. 9 (Q[0] = "1"), the single-ended writing has less WM, which leads to an easier write failure. To address this, VCC2 is boosted beyond VDD, which raises the switching threshold of INV II. This makes data flipping easier. If the write data is "1," as shown in Fig. 8(b), then WBLB[0] is connected to VDD, and VCC1[0] is boosted. During the write operation, the unselected cells in the selected column are similar to the half-selected cells in the conventional SRAMs. In the unselected rows, WR_VGNDs are floating when writing data. If the data stored in the half-selected cell is "0" while writing "0," then WBL[0] will be set to VDD [as Q_{10} in Fig. 8(b)]. Since the unselected WR_VGNDs are precharged to high, the current flows from WR_VGNDs to Q_{x0} ($x \in (1, (m-1))$), and the voltage of Q_{x0} goes up.

TABLE III
COMPARISON OF READ-ASSIST WITH PREVIOUS WORK

Assist types	Conv. 8T	[25]	[26]	[24]	This Work
schematic					
Cell Size*1	1.3x	1.3x	1.3x	1.15x	1.3x
Read disturb	No	No	No	No	No
Discharged RBLs*2	8	8	16	8	1
Leakage *3	Yes	No	No	No	No
RBL swing					

*1. This table is only discussing the read port. The write port just assumes all as conventional 6T SRAM. All area information in the table is based on the assumption.
*2. This number is based on 50% '1' and 50% '0' data pattern in the selected row in one subarray with 16 columns which is using one CMUX.
*3. The leakage is in the hold period, only for the leakage between two sides of the read port.

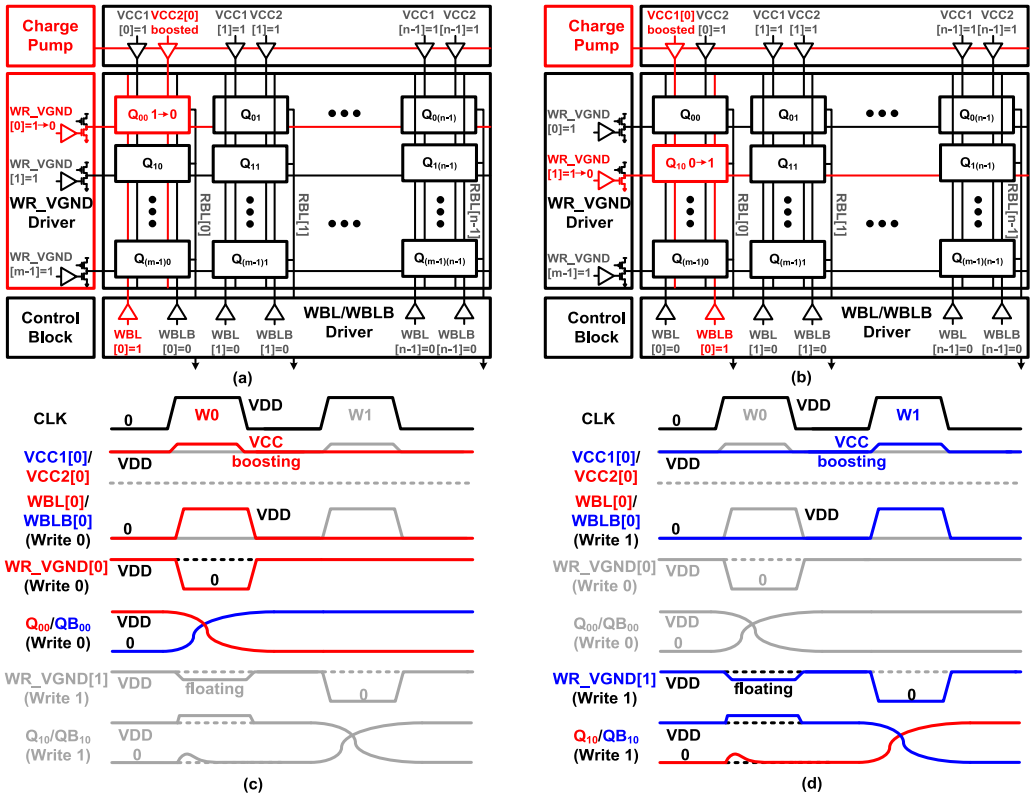


Fig. 8. Principle of the proposed write operation. (a) Topology of writing "0" operation. (b) Topology of writing "1" operation. (c) Timing diagram of writing "0" operation. (d) Timing diagram of writing "1" operation.

The data in Q_{x0} and QB_{x0} could be flipped because of the disturbing current. However, the proposed write technique utilizes boosted VCC2 to overdrive the pull-down transistor on the Q_{x0} side. This discharges the disturbing current on Q_{x0} quickly. As shown in Fig. 9 (middle), if Q[m] reaches the switching threshold voltage of INV II, data will flip. In this case, V_{st2} could be improved by boosting VCC2, avoiding the data flips in the half-selected condition. If the data stored in the half-selected cell is "0" while writing "1," then WBLB[0] will be asserted [as QB_{00} in Fig. 8(a)], the unselected WR_VGNDs are precharged to logic "1" and then floating. No disturbing current flows from QB_{00} to WR_VGND during the write operation. Even with boosted VCC1[0], the pull-up transistor on the Q_{x0} side is open since QB_{x0} is "1." Therefore, the voltage at Q_{x0} is not affected by the boosted VCC1[0]. As shown in Fig. 9 (bottom), it is a mirrored case for QB_{00}. If we boost VCC2, because Q[n] is 1, then M2 is turned off, the value of QB[n] maintains at GND. In order to minimize the write time and avoid the fluctuations on the boosted signal at the beginning, WBL/WBLB is asserted after VCC1/2 is

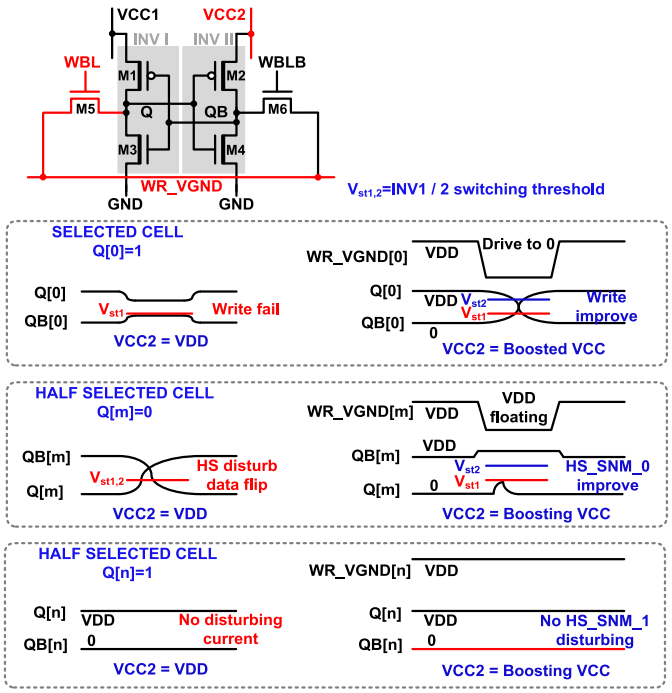


Fig. 9. Impact of data-aware split power lines to writing "0" operation in selected and half-selected cells.

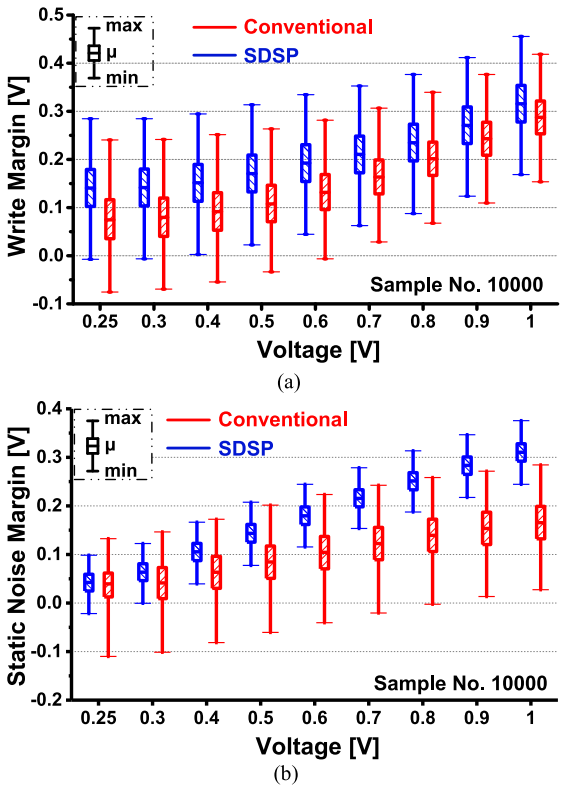


Fig. 10. Comparison of (a) WM and (b) SNM. The simulation was performed using the SRAM cell schematics.

applied. Consequently, the data-aware SDSP scheme improves the write ability and strengthens the stability simultaneously. For the hold period, both VCC1 and VCC2 are at VDD, the WBL/WBLB is "0" to disable the pass gate transistors, and all of WR_VGND are precharged to VDD. Since the surrounding nodes are held at the same potential, the overall leakage current is similar to the conventional one implemented with the same size and the same threshold devices.

Fig. 10 presents the statistical analysis of WM and SNM from schematic-level simulation results. The number of sample points is 10 000, the temperature is 27 °C, and VCC1/2 is boosted to 125% of VDD. In WM [Fig. 10(a)], the convention starts to fail when the supply voltage drops to 0.6 V. At 0.3 and 0.25 V, the minimum WMs become −69.4 and −75.4 mV, respectively. The proposed SDSP scheme starts to fail from 0.3 V, with a bit error rate of 0.01% out of 10 000 samples. When the voltage scales to 0.25 V, the write fail-rate increases to 0.24%, and the minimum WM becomes −7.4 mV. Fig. 10(b) presents the stability of the conventional 8T SRAM and the proposed SRAM. The conventional SRAM begins to fail when the supply voltage is 0.8 V. The worst cases of SNM at 0.3 and 0.25 V are −101 and −110 mV, respectively. However, the proposed SRAM has no failure until the supply voltage drops below 0.3 V. At 0.25 V, the failure rate of 0.71% is achieved. Overall, the proposed SDSP scheme improves both WM and SNM simultaneously.

B. VSSUP Versus VCC_BST for Stability

In the conventional SRAM cell, stability and write ability have contradictory requirements in the write operation. Stronger write access transistors lead to larger WM, whereas they worsen SNM. However, in the proposed SRAM, both data-aware split ground lines VSS1, 2 (raising VSS value signed as VSSUP) [12], [13] and power lines VCC1, 2 (boosting the VCC value signed as VCC_BST) could improve WM and half-selected static noise margin (HS-SNM) simultaneously. Fig. 11 presents the simulated SNM using VSSUP and VCC_BST. As shown in Fig. 11(a), the half-selected cell SNM for data storing "0" (HS-SNM-0) is improved by raising VSS, while the half-selected cell SNM for data storing "1" (HS-SNM-1) degrades. However, since HS-SNM-1 is much larger than HS-SNM-0, the worst case SNM is still improved. Note that if VSSUP is raised beyond 20% of VDD in this article, HS-SNM-1 becomes dominant and the worst case SNM starts to drop. For the VCC_BST scheme, as shown in Fig. 11(b), HS-SNM-0 also improves with boosting of VCC, whereas HS-SNM-1 decreases slowly compared to the VSSUP scheme. In this case, HS-SNM-0 is still limiting when VCC is boosted by 20%. Fig. 11(c) compares the impacts of the VSSUP and VCC_BST schemes on stability. Since VCC_BST shows better stability improvement, we utilized VCC_BST in this article.

Each power lines (VCC1 or VCC2) drives 256 cells in one column, and 16 columns share one VCC_BST generator. Fig. 12(a) shows the schematic in the column of the implemented boosted voltage generator. VCC1 and VCC2 are controlled by the data input D_{in} and write column signal WR_Col. S0 and S1 are controlling the boosted voltage generators, and the voltage level of S0/S1 is the same as the supply voltage. S0 is used during all write operations, and S1 is activated for ultralow voltage only (below 0.4 V) to increase the capacitor for 20% boosted voltage generating.

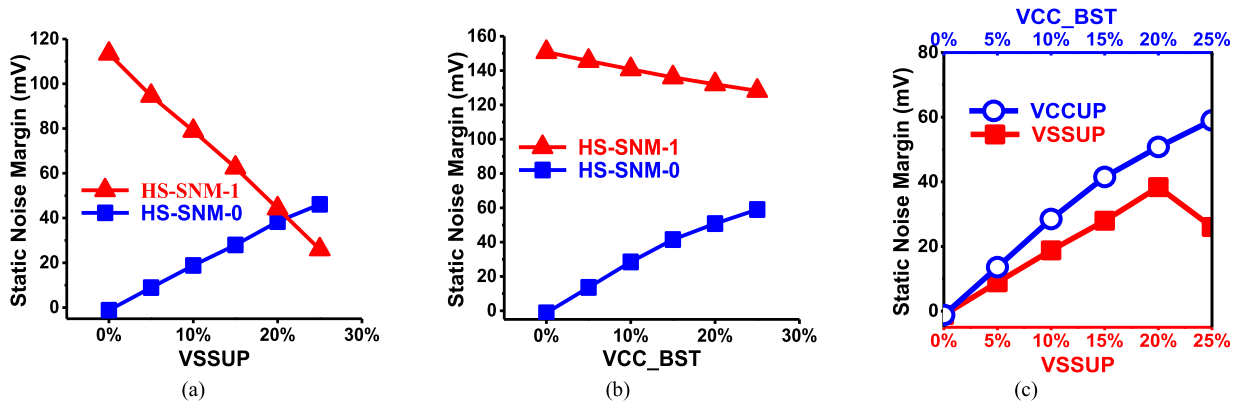


Fig. 11. (a) Impact of VSSUP to SNM under 0.4 V. (b) Impact of VCC_BST to SNM under 0.4 V. (c) Impact comparison of VCCUP and VSSUP on SNM.

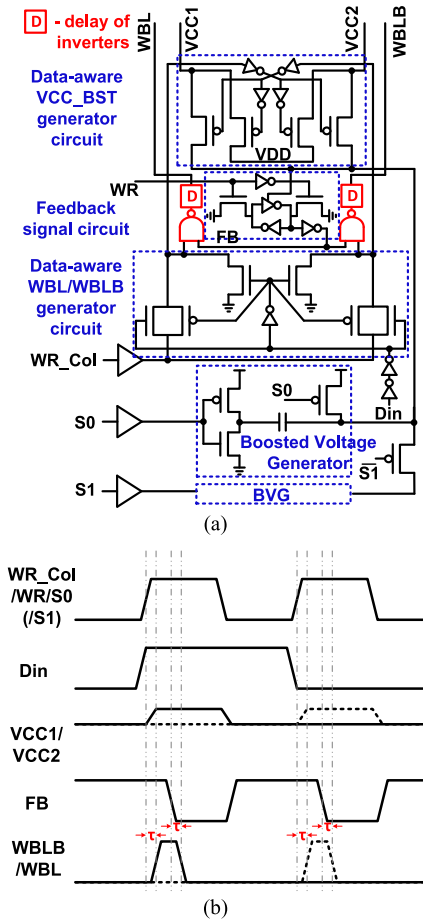


Fig. 12. (a) Circuit in column to generate VCC_BST. (b) Time diagram of the VCC_BST generating circuit.

Note that either VCC1 or VCC2 is connected to the VCC_BST node during the write operation. If not, WBL and WBLB are set to VDD, driving VCC1 and VCC2 with VDD. Fig. 12(b) shows the timing diagram of the boosted voltage generator. The delay (τ) between WR_Col and WBL/WBLB is decided by the delay of the NAND and the delay of inverters, which is the same as the delay between feedback (FB) and the falling edge of WBL/WBLB.

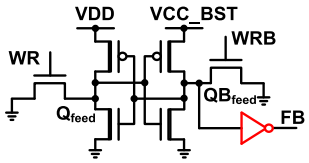


Fig. 13. Feedback signal generator circuit.

C. Write-Completion Flag Generation

This technique improves WM and the write speed at the cost of energy overhead. To minimize the energy overhead, this work employs a write-completion flag generator for resetting WBL/WBLB once a write operation is completed. A modified SRAM cell is duplicated in each column as shown in Fig. 13. The write access transistors are controlled by the write enable signals (WR and WRB). During the write operation, WR is enabled, the node Q_{feed} is discharged. Once the state of QB_{feed} is flipped, the write-completion flag signal (FB) will disable the asserted WBL/WBLB to reduce the redundant energy consumption. Since QB_{feed} is connected to an inverter, QB_{feed} is flipped later than the cells in the array due to the parasitic capacitance. In addition, extra inverters can be added (red part in Fig. 13) to increase the delay. Overall, the parasitic capacitor and the inverters provide enough timing margin for FB. Consequently, the write-completion flag minimizes the energy overhead by disabling WR and WRB as quickly as possible.

D. Architecture of 16-kb SRAM

Fig. 14 shows the architecture of the proposed 16-kb SRAM. The test chip is configured of 4 × 4 kb SRAM macros, with each 4-kb SRAM macro is implemented with a 256 × 16 array, write word line (WWL)_VGND, RWL_VGND and column decoder array, data-aware VCC_BST switch circuit, local amplifier, feedback loop, and boosted voltage generator. WR_VGND (RD_VGND) is driven by a WWL_VGND (RD_VGND) decoder for four switch circuits, each one is for 16 columns. VCC1s and VCC2s are generated from a column decoder using write data. A boosted voltage generator is placed at the middle and bottom of each macro with the data-in and

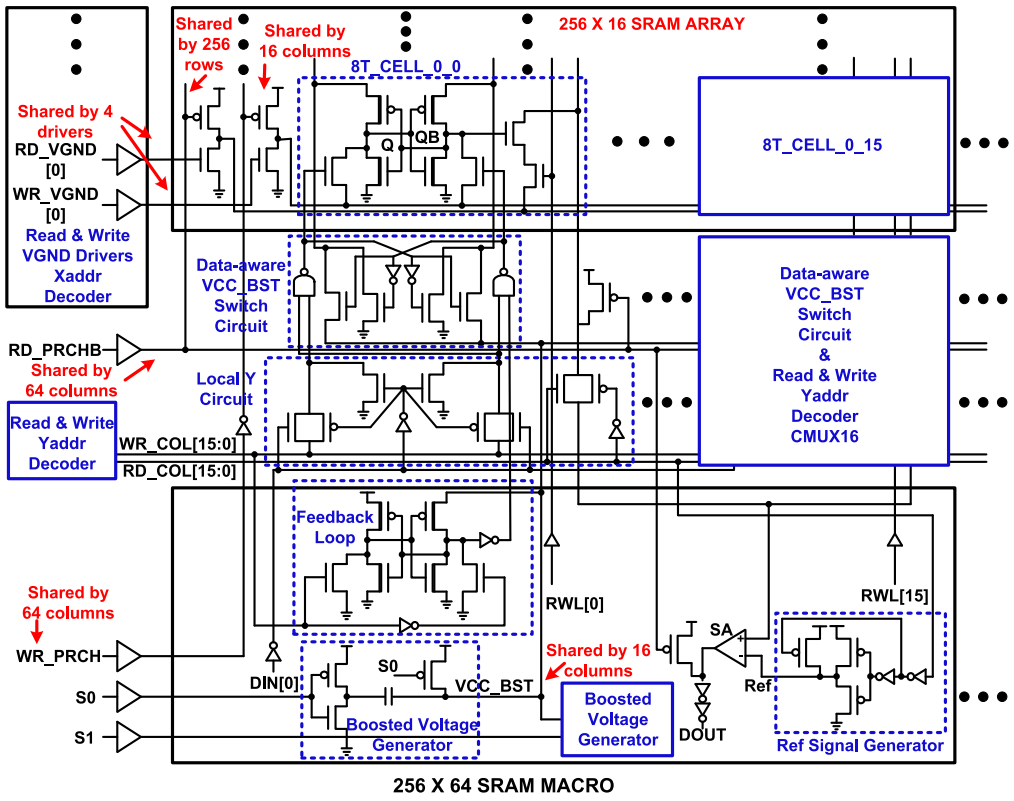
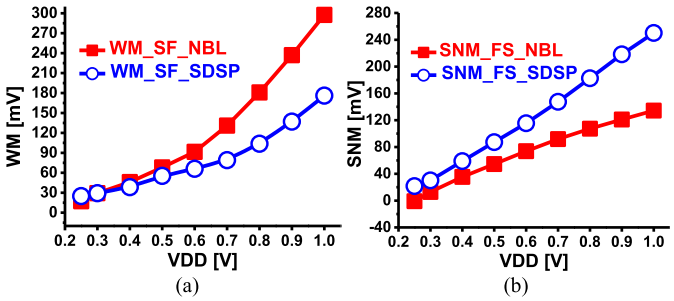
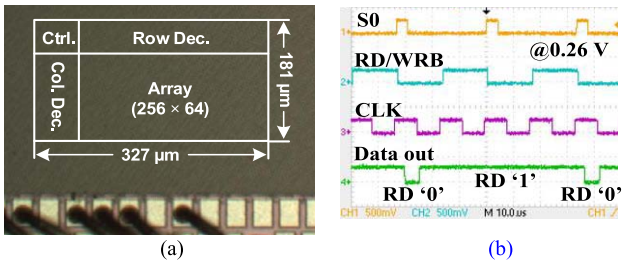


Fig. 14. Schematic with the proposed assist circuit.



(a) (b)

Fig. 15. Comparison of WM and SNM between NBL and SDSP.



(a) (b)

Fig. 16. (a) Die photograph of the proposed 16-kb SRAM. (b) Waveform of test chip at 0.26-V voltage supply.

data-out buffers. The proposed write-assist technique creates an area overhead of 1.37% in the implemented 16-kb SRAM chip. It also increases power and leakage by 7.9% and 2%, respectively.

E. Write-Assists Comparison

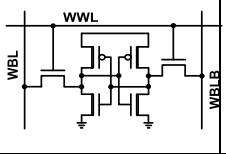
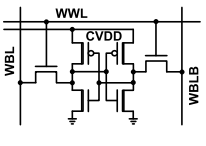
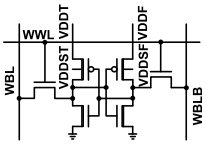
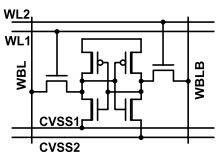
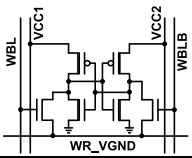
NBL is one of the widely used write-assist techniques [10], [13], [14]. Fig. 15 compares WM and SNM in SDSP and NBL. As shown in Fig. 15(a), NBL has better WM in the super-threshold region and shows similar WM at the target operation point (i.e., 0.4 V). However, SDSP shows a larger SNM than NBL in the whole voltage range. Therefore, it can be said that SDSP is more practical compared to NBL. In terms of power, SDSP consumes more power than NBL in the selected column due to the larger number of devices connected to VCC1/VCC2. However, no dynamic power is consumed in the unselected columns. Our analysis shows that SDSP and NBL dissipate similar power when using a 16:1 multiplexing ratio.

Table IV compares the proposed SDSP scheme with other state-of-the-art write-assist techniques. For a fair comparison, all of the write-assist schemes are implemented and simulated in the employed technology node. It can be seen that other write-assist techniques improve WM while degrading the HS-SNM. However, the proposed SDSP technique improves both WM and HS-SNM simultaneously.

IV. Measurement Results

We fabricated a 16-kb SRAM in 65-nm CMOS technology. It comprises four subblocks (each block consists of 256 rows × 16 columns) and generates four-bit data for test purposes. Fig. 16(a) shows the test chip micrograph. It occupies 327 × 181 μm^2 area. The sample-measured waveforms in Fig. 16(b) demonstrate the successful operation

TABLE IV
COMPARISON OF WRITE-ASSIST WITH PREVIOUS WORK

Assist types	Conv. 6T	DTWL*[1] + NBL [14]	DDDS*[1]+BM*[1] [18]	DSC*[1]+NBL [10]	DSC*[1]+DSPL This Work
schematic					
Cell Size	1x	1x	1.75x	1x	1x
HS assist	No	DTWL	FBB	S-WL + S-VSS	S-WL + S-VCC
Write assist	No	DTWL + NBL	DDDS	S-VSS + NBL	S-WL + S-VCC
Leakage	NI*[2]	NI	Degrades	NI	NI
WM*[3,4]	Fail	x0.63	x2.48	x2.5	x1
HS-SNM*[3,4]	Fail	x0.23	x0.34	x0.34	x1

*1 DTWL: Dual transient wordline; DDDS: Data dependent differential supply; BM: Body modulation; FBB: Forward body bias; DSC: Dual split control.
*2 NI: No/negligible impact.
*3 Data is under 0.5V supply voltage, 27°C, WM is in SF corner, SNM is in FS corner.
*4 ΔNBL=0.2*VDD, ΔWLUD=0.2*VDD, ΔWLOD=0.2*VDD, ΔDDDS=0.2*VDD, ΔVSS=0.2*VDD, ΔVCC=0.2*VDD.

TABLE V
COMPARISON OF SRAM PERFORMANCE

	JSSC 2013[17]	TCAS-I 2016[6]	JSSC 2017[10]	JSSC 2017[14]	JSSC 2018[27]	This Work
Technology	65nm	65nm	28nm	28nm	65nm	65nm
Density	2kb	16kb	256kb	32kb	2kb	16kb
Transistor count	9T	9T	6T	8T	6T	8T
Cell area	675.03 F^2	448.19 F^2	161.99F^2	708.56F^2	508.88 F^2	538.37 F^2
VDDmin	0.28V	0.26V	0.5V	0.358V	0.29V	0.26V
Frequency	330kHz (0.32V)	N.A.	20MHz (0.5V)	50MHz	3.34MHz	1.4MHz (0.4V)
Leakage current	0.05μA (0.4V)	1.4μA (0.1V)	120μA (0.5V)	N.A.	8.7uA (0.29V)	0.22μA (0.26V)
Normalized Leakage	24.4pA/b (0.4V)	85.44pA/b (0.1V)	457.8pA/b (0.5V)	N.A.	4.25nA/b (0.29V)	14pA/b (0.26V)
Min. energy	0.57pJ	2.07pJ	30pJ	N.A.	2.25pJ (0.4V)	0.506pJ (0.4V)
Normalized energy	278aJ/b	126aJ/b	120aJ/b	N.A.	1.1fJ/b	31aJ/b

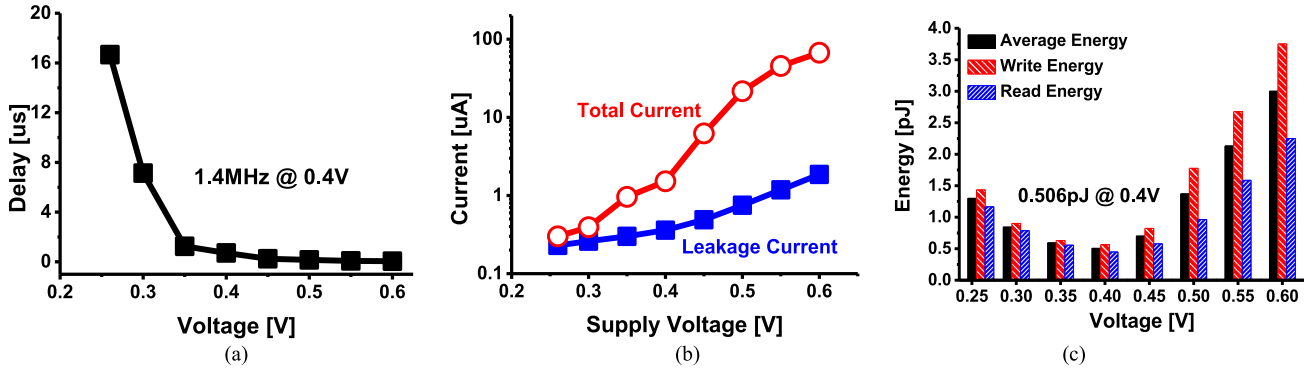


Fig. 17. Measured results. (a) Minimum clock period. (b) Leakage current and total current. (c) Energy consumption.

of the proposed SRAM at 0.26 V. At 0.26 V, our fabricated chip achieved a maximum frequency and energy consumption of 60 kHz and 78 nW, respectively [shown in Fig. 17(a)]. Fig. 17(b) shows the leakage current and the total current at room temperature. The leakage was measured in the standby mode after disabling the clock signal. It is reduced significantly because of the virtual ground, eliminating the subthreshold leakage in the read port. At ultralow voltage (under 0.35 V), the leakage current becomes dominant. The measured minimum energy dissipation is 0.506 pJ at 0.4 V [Fig. 17(c)], while the operation frequency achieves 1.4 MHz. Note that the total current and the energy were measured using balanced read and write operations with balanced data patterns. The proposed 8T SRAM is compared with other state-of-the-art works in Table V. The proposed vertical RWL reduces the dynamic energy significantly without substantial performance degradation. Consequently, the proposed SRAM achieved the minimum energy of 31 aJ/b at 0.4 V.

V. Conclusion

In this article, a 16-kb 8T SRAM array with vertical RWL and SDSP for ultralow-power applications has been presented. The column-based RWL reduces the dynamic energy consumption by only discharging the selected RBL in one column MUX (CMUX). A single-ended write operation is developed with vertical control. Finally, the proposed data-aware SDSP enhances WM and SNM simultaneously. The test chip of the proposed SRAM is fabricated in 65-nm CMOS technology. Based on the experimental result, the minimum energy per bit of 31 aJ is achieved at 0.4 V. This achieves the smallest normalized energy per bit when compared to other state-of-the-art works. The SRAM is successfully functional down to 0.26 V.

References

[1] E. J. Nowak *et al.*, "Turning silicon on its edge, overcoming silicon scaling barriers with double-gate and FinFET technology," *IEEE Circuit Device Mag.*, vol. 20, no. 1, pp. 20–31, Jan./Feb. 2004.

[2] C. T. Chuang, S. Mukhopadhyay, J.-J. Kim, K. Kim, and R. Rao, "High-performance SRAM in nanoscale CMOS: Design challenges and techniques," in *Proc. IEEE Int. Workshop Memory Technol., Design Test.*, Dec. 2007, pp. 4–12.

[3] H. Jeong *et al.*, "Offset-compensated cross-coupled PFET bit-line conditioning and selective negative bit-line write assist for high-density low-power SRAM," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 4, pp. 1062–1070, Apr. 2015.

[4] W. Choi, G. Kang, and J. Park, "A refresh-less eDRAM macro with embedded voltage reference and selective read for an area and power efficient Viterbi decoder," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2451–2462, Oct. 2015.

[5] I. J. Chang, J.-J. Kim, S. P. Park, and K. Roy, "A 32 kb 10 T sub-threshold SRAM array with bit-interleaving and differential read scheme in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 2, pp. 650–658, Feb. 2009.

[6] B. Wang, T. Q. Nguyen, A. T. Do, J. Zhou, M. Je, and T. T. H. Kim, "Design of an ultra-low voltage 9 T SRAM with equalized bitline leakage and CAM-assisted energy efficiency improvement," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 2, pp. 441–448, Feb. 2015.

[7] M. H. Tu *et al.*, "A single-ended disturb-free 9 T subthreshold SRAM with cross-point data-aware write word-line structure, negative bit-line, and adaptive read operation timing tracing," *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1469–1482, Jun. 2012.

[8] L. Wen, X. Cheng, K. Zhou, S. Tian, and X. Zeng, "Bit-interleaving-enabled 8 T SRAM with shared data-aware write and reference-based sense amplifier," *IEEE Trans. Circuit Syst. II, Exp. Briefs*, vol. 63, no. 7, pp. 643–647, Jul. 2016.

[9] Y. Wang *et al.*, "A 1.1 GHz 12 μA/Mb-leakage SRAM design in 65 nm ultra-low-power CMOS technology with integrated leakage reduction for mobile applications," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 172–179, Jan. 2008.

[10] S. L. Wu *et al.*, "A 0.5-V 28-nm 256-kb mini-array based 6 T SRAM with Vtrip-tracking write-assist," *IEEE J. Solid-State Circuits*, vol. 64, no. 7, pp. 1791–1802, Mar. 2017.

[11] T. Song *et al.*, "A 14 nm FinFET 128 Mb SRAM with V_{MIN} enhancement techniques for low-power applications," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 158–169, Jan. 2015.

[12] O. Hirabayashi *et al.*, "A process-variation-tolerant dual-power-supply SRAM with 0.179 μm^2 cell in 40 nm CMOS using level-programmable wordline driver," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2009, pp. 458–459.

[13] M. F. Chang *et al.*, "A compact-area low-VDDmin 6 T SRAM with improvement in cell stability, read speed, and write margin using a dual-split-control-assist scheme," *IEEE J. Solid-State Circuits*, vol. 52, no. 9, pp. 2498–2514, Sep. 2017.

[14] T. Song *et al.*, "A 10 nm FinFET 128 Mb SRAM with assist adjustment system for power, performance, and area optimization," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 240–249, Jan. 2017.

[15] Y. H. Chen *et al.*, "A 16 nm 128 Mb SRAM in high-κ metal-gate FinFET technology with write-assist circuitry for low-VMIN applications," *IEEE J. Solid State Circuits*, vol. 50, no. 1, pp. 170–177, Jan. 2015.

[16] G. Pasandi *et al.*, "A 256-kb 9 T near-threshold SRAM with 1 k cells per bitline and enhanced write and read operations," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 23, no. 11, pp. 319–332, Nov. 2015.

[17] S. Lutkemeier, T. Jungeblut, H. K. O. Berge, S. Aunet, M. Porrmann, and U. Ruckert, "A 65 nm 32 b subthreshold processor with 9 T multi-Vt SRAM and adaptive supply voltage control," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 2438–2446, Jan. 2013.

[18] A. Grover *et al.*, "A 32 kb 0.35–1.2 V, 50 MHz–2.5 GHz bit-interleaved SRAM with 8 T SRAM cell and data dependent write assist in 28-nm UTBB-FDSOI CMOS," *IEEE J. Solid-State Circuits*, vol. 64, no. 9, pp. 2438–2447, Sep. 2017.

[19] L. Lu, T. Yoo, and T. T.-H. Kim, "An ultra-low power 8 T SRAM with vertical read word line and data aware write assist," in *Proc. Asian Solid-State Circuits Conf. (A-SSCC)*, Nov. 2018, pp. 143–144.

[20] L. Chang, M. Ieong, and M. Yang, "CMOS circuit performance enhancement by surface orientation optimization," *IEEE Trans. Electron Devices*, vol. 51, no. 10, pp. 1621–1627, Sep. 2004.

[21] F. Hamzaoglu *et al.*, "A 3.8 GHz 153 Mb SRAM design with dynamic stability enhancement and leakage reduction in 45 nm high-k metal gate CMOS technology," *IEEE J. Solid State Circuit*, vol. 44, no. 1, pp. 148–154, Jan. 2009.

[22] I. Fukushi *et al.*, "A system level memory power optimization technique using multiple supply and threshold voltages," in *IEEE Symp. VLSI Circuits. Dig. Tech. Papers*, Jan. 1998, pp. 142–145.

[23] S. Gupta, K. Gupta, B. H. Calhoun, and N. Pandey, "Low-power near-threshold 10 T SRAM bit cells with enhanced data-independent read port leakage for array augmentation in 32-nm CMOS," *IEEE Trans. Circuit Syst. I, Reg. Papers*, vol. 66, no. 3, pp. 978–988, Mar. 2019.

[24] M.-F. Chang *et al.*, "A sub-0.3 V area-efficient L-shaped 7 T SRAM with read bitline swing expansion schemes based on boosted read-bitline, asymmetric-Vth read-port, and offset cell VDD biasing techniques," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2558–2569, Oct. 2013.

[25] D. Takashima *et al.*, "A 7 T-SRAM with data-write technique by capacitive coupling," *IEEE J. Solid-State Circuits*, vol. 54, no. 2, pp. 596–605, Feb. 2019.

[26] C.-F. Chen, T.-H. Chang, L.-F. Chen, M.-F. Chang, and H. Yamauchi, "A 210 mV 7.3 MHz 8T SRAM with dual data-aware write-assists and negative read wordline for high cell-stability, speed and area-efficiency," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2013, pp. 130–131.

[27] M. Nabavi and M. Sachdev, "A 290-mV, 3.34-MHz, 6 T SRAM with pMOS access transistors and boosted wordline in 65-nm CMOS technology," *IEEE J. Solid-State Circuits*, vol. 53, no. 2, pp. 656–667, Feb. 2018.

Lu Lu (S'17) received the B.E. degree from the School of Computer and Information, Hefei University of Technology, Hefei, China, in 2007, and the M.E. degree from the School of Microelectronics and Solid State Electronics, Xiamen University, Xiamen, China, in 2010. She is currently working toward the Ph.D. degree at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

Her research interests include low-power static random access memory (SRAM) and SRAM-based physical unclonable function (PUF).

Ms. Lu was a recipient of the IEEE SSCS Singapore Chapter Award in 2018.

Taegeun Yoo (S'09–M'17) received the B.S., M.S., and Ph.D. degrees in electrical and electronics engineering from Chung-Ang University, Seoul, South Korea, in 2009, 2011, and 2015, respectively.

From 2015 to 2016, he was a Research Professor with Chung-Ang University. In 2016, he joined as a Research Fellow with Nanyang Technological University, Singapore. His research interests include analog-mixed signal ICs and low-power memory architecture.

Dr. Yoo received encouragement award and silver award at the Human-Tech Paper Award hosted by Samsung Electronics, in 2011 and 2014, respectively, and also received the Silkroad Award at the IEEE International Solid State Circuits Conference (ISSCC) in 2014.

Van Loi Le (S'16) received the B.E. degree in electronics and telecommunication engineering from the University of Science and Technology, The University of Da Nang, Da Nang, Vietnam, in 2014, and the Ph.D. degree in electrical and electronics engineering from Nanyang Technological University, Singapore, in 2019.

His research interests include energy-efficient digital signal processor design, ultralow power, and ultralow voltage digital circuit design.

Dr. Le was a recipient of the Joint Industry Postgraduate Program Scholarship in 2015, the Patent Award from NXP Semiconductors in 2016, the IEEE SSCS Singapore Chapter Award in 2018, the Best Poster Award KSEASG Symposium for Young Generation Scientists and Engineers in 2018, the third prize at EEE Graduates' Research Showcase in 2018, IEEE Solid-State Circuits Society Student Travel Grant Award in 2019, and the Award of Excellence from Nations Technologies in 2019.

Tony Tae-Hyoung Kim (M'06–SM'14) received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, South Korea, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2009.

From 2001 to 2005, he was with the Samsung Electronics, Hwasung, South Korea, where he performed research on the design of high-speed static random access memories (SRAMs), clock generators, and IO interface circuits. From 2007 to 2009, he was with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, and Broadcom Corporation, Edina, MN, USA, where he performed research on circuit reliability, low-power SRAM, and battery-backed memory design. In 2009, he joined Nanyang Technological University, Singapore, where he is currently an Associate Professor. He has authored or coauthored over 160 journal and conference articles. He holds 17 U.S. and Korean registered patents. His current research interests include low-power and high-performance digital, mixed-mode, and memory circuit design, ultralow-voltage circuits and systems design, variation and aging-tolerant circuits and systems, and circuit techniques for 3-D ICs.

Dr. Kim received the Best Demo Award at APCCAS2016; the Low Power Design Contest Award at ISLPED2016; the Best Paper Awards at 2011 ISOCC and 2014 ISOCC; the AMD/CICC Student Scholarship Award at the IEEE CICC2008; the Departmental Research Fellowship from the University of Minnesota in 2008; the DAC/International Solid State Circuits Conference (ISSCC) Student Design Contest Award in 2008; the Samsung Humantec Thesis Award in 2008, 2001, and 1999; and the ETRI Journal Paper of the Year Award in 2005. He was the Chair of the IEEE Solid-State Circuits Society Singapore Chapter. He has served as a committee member for numerous conferences. He serves as an Associate Editor for the IEEE TRANSACTIONS ON VERY-LARGE-SCALE INTEGRATION (VLSI) SYSTEMS, IEEE ACCESS, and the *IEIE Journal of Semiconductor Technology and Science*.