A CAM-ASSISTED 9T SRAM CIRCUIT FOR AN ULTRA-LOW VOLTAGE OPERATION

GUIDE
Dr.P.Ramesh Kumar,M.Tech.,Ph.D.,

Professor,

Department of ECE

BY
M.Kalaiarasi,

II YEAR M.E.,

VLSI DESIGN,

Department of ECE

SENGUNTHAR ENGINEERING COLLEGE

Abstract—**This paper presents a 9T multi-threshold (MTCMOS) SRAM macro with equalized bitline leakage and a content-addressable-memory-assisted (CAM-assisted) write performance boosting technique for energy efficiency improvement. A 3T-basedread port is proposed to equalize read bitline (RBL) leakage and**

to improve RBL sensing margin by eliminating data-dependence on bitline leakage current. A miniature CAM-assisted circuit is integrated to conceal the slow data development with HVT devices after data flipping in write operation and therefore enhance the write performance for energy efficiency. A 16 kb SRAM test chip

is fabricated in 65 nm CMOS technology. The operating voltage of the test chip is scalable from 1.2 V down to 0.26 V with the read access time from 6 ns to 0.85 us . Minimum energy of 2.07pJ is achieved at 0.4 V with 40.3% improvement compared to the SRAM without the aid of the CAM. Energy efficiency is enhanced by 29.4% between 0.38 V 0.6 V by the proposed CAM-assisted circuit.

Index Terms—**Bitline leakage equalization, content addressablememory, energy efficiency improvement, ultra-low voltage SRAMdesign.**

I. INTRODUCTION

STATE-OF-THE-ART DSP cores and advanced healthcaressoCs ,benefit from availability of on-chip SRAMswith substantially reduced power dissipation and improved energyefficiency. Integrated SRAMs play a crucial role in providing the required density, performance, power, and energy consumption of applications. By aggressively scaling supply voltage near or below transistor's threshold voltage, power and energy efficiency of SRAMs can be greatly ameliorated at the expense of performance. However, the vulnerability of SRAMsto PVT fluctuations makes reliable near- and sub-threshold operations extremely challenging in deep sub-micrometer

CMOS technologies. Simultaneously, other design metrics such as stability, read/write margin, and leakage need to be carefully revisited for the reliable operation. SRAMs have achieved ultra-low power/energy through supply scaling [3]–[5]. However, they suffer from various design issues mainly caused by reduced Ion -to- Ioff ratio combined with large variations. Under severely scaled supply voltage, cell stability and bitline sensing margin of 6T SRAMs degrade dramatically due to the significant impact of disturbing current and bitline leakage. To handle it, an 8T differential SRAM cell has been proposed to inject identical leakage current into the differential bitlines, eliminating the differential offset voltage from the leakage. However, in general, decoupled SRAM cells are preferable in weak-inversion regime to make the read static-noise-margin (SNM) identical to the hold SNM. Moreover, the dedicated read port enables a faster read operation with no disturbing current to cell nodes. Energy efficiency is a vital design metric for ultra-low voltage SRAMs. Although voltage scaling decreases the switching energy quadratically, it deteriorates the operating frequency by several orders of magnitude. Accordingly, leakage energy accumulated in slow clock cycles would dominate the total energy in the deep sub-threshold region, leading to an energy contour shooting up.To reduce the static energy, leakage current minimization techniques are desirable. In general logic circuits, adoption of HVT devices in non-critical paths is favorable to suppress the leakage. Another effective method to improve energy efficiency is suppressing leakage energy by eliminating idle gates or modules in the system, which is adopted by.

Leakage suppression is also attainable from algorithm level. Among all the strategies for energy saving, leakage energy reduction is the first concern to improve energy efficiency. In this work, we present several design techniques to foster an energy efficient SRAM in a wide range of supply voltages with the following features: i) a decoupled 9T SRAM cell with an improved SNM compared to the 6T cell;

ii) a 3T read port for equalizing RBL leakage and augmenting bitline swing; iii) utilizing MTCMOS technology for minimizing leakage in 6T write part and maximizing SRAM performance in read port; iv) a CAM-assisted circuit technique for improving the energy efficiency by boosting the write speed. The proposed circuit techniques are demonstrated by a 16 kb SRAM test macro (including

the CAM) fabricated in a 65 nm CMOS technology.

B. Wang, T. Q. Nguyen, A. T. Do, and T. T.-H. Kim are with the Schoolof Electrical and Electronic Engineering, Nanyang Technological University,639798 Singapore (e-mail: bwang6@e.ntu.edu.sg; quynh2741@hotmail.com; atdo@ntu.edu.sg; thkim@ntu.edu.sg).
J. Zhou is with the Institute of Microelectronics, A*STAR, 117685 Singapore(e-mail: zhouj@ime.a-star.edu.sg).

M. Je was with the Institute of Microelectronics, A*STAR, 117685 Singapore.He is now with the Department of Information and Communication Engineering,Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu711-873, Korea (e-mail: Minkyu.je@dgist.ac.kr).
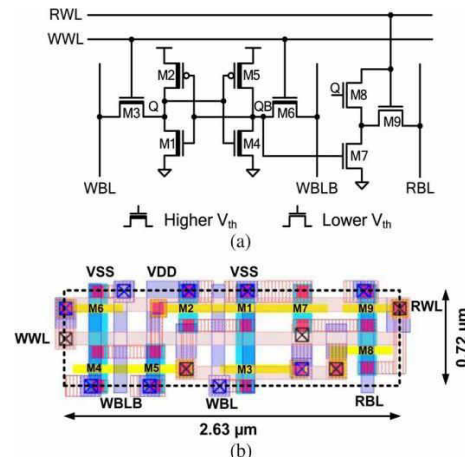
Fig. 1. (a) Proposed 9T SRAM cell implemented with HVT devices in write paths and LVT devices in read port.   (b) Layout of the 9T SRAM cell based on 65 nm logic design rules.

II. PROPOSED SRAM DESIGN TECHNIQUES FOR ULTRA-LOW VOLTAGE OPERATION

A. Proposed 9T SRAM Cell

Fig. 1 depicts the proposed 9T SRAM cell and its layout. The cell consists of a 6T SRAM part (the write-access transistors with a cross-coupled latch) and a dedicated read port. The read port comprises three NMOS transistors (M7, M8, and M9) for realizing equalized bitline leakage and improving bitline sensing margin in a single-ended read bitline (RBL). The write access paths and the data storage latch are implemented with HVT devices for leakage reduction while the read port employs LVT devices for performance. The layout of the 9T

cell occupies an area of 2.63 0.72 based on logic design rules. A write operation is enabled by activating a write wordline (WWL) and completed when the data loaded at WBL and

WBLB is written into Q and QB. A read operation starts by enabling a read wordline (RWL) and is followed by conditional

RBL discharging. If Q holds logic 0, M7 is turned on and discharges RBL to GND. If, on the contrary, Q stores logic 1, M8 is activated and provides pull-up current from RWL (=V) to

RBL, slowing down the discharging speed of RBL.

B. Analysis of Static Noise Margin and Write Margin

Decoupled SRAM cells, such as the 8T SRAM cell in and the 10T SRAM cell in [10], have been widely accepted for SNM improvement. Eliminating the interference from read bitlines

into cell nodes, such as the 8T cell and the 10T cell, makes the read-mode SNM equivalent to the hold-mode SNM. The read-mode SNM of the proposed 9T

multi-V SRAM cell is compared to those of the conventional 6T cell and the 10T cell in Fig. 2(a). To investigate the impact of different V on SNM, the 6T SRAM cell is implemented with two device types. One

is implemented with SVT devices and the other is implemented with HVT devices. Both pull-down NMOS transistors are oversized

by 1.67 . SVT devices with the same geometry as the 9T SRAM cell are utilized for the 10T cell with the assumption that no multi-threshold voltage option is adopted in. The SNM values over the operating supply range are illustrated

in Fig. 2(a). For the SVT cells, SNMs increase significantly with VDD and then slightly slows
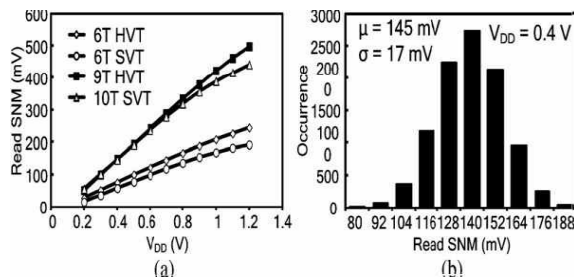

Fig. 2. (a) 9T read SNMs compared with 6T and 10T SNMs at different voltages. (b) Distribution of 9T read SNMs at 0.4 V supply voltage.
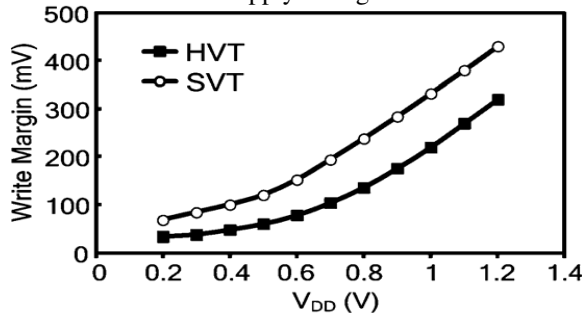

Fig. 3. Comparison of write margins on HVT and SVT devices. down in the super-threshold regime. The SNMs of the 6T and the 9T HVT cells, whereas, exhibit a more linear slope with supply voltage, which are far from saturation with increased V . This is partially caused by higher channel implant by HVT layer in this multi-threshold

technology. The 9T cell shows a SNM of 52 mV at 0.2 V, improving the margin by 85.7% compared to the 6T HVT cell. At the nominal supply voltage, the SNM of the 9T SRAM cell is

1.13 larger than the 10T SRAM cell whereas the difference of the two SNMs decreases at lower supply voltages. SNM Monte Carlo simulations VDD=0.4Vfor 3 mismatch on top of the TT corner are conducted and the results are illustrated in Fig. 2(b). The 10k-point Monte Carlo simulations at reveal that the proposed SRAM cell generates a mean

SNM of 145 mV with a standard deviation of 17 mV. It provides a higher mean value with comparable variation than the 10T SRAM cell composed of all standard V (SVT) transistors. For an SRAM cell, write margin is interpreted as the voltage

headroom at write wordline for a successful write operation. Generally, it is determined by the drive strength ratio of the write-access transistors to the pull-up transistors. Simulated

write margin of the 9T SRAM cell is plotted in Fig. 3. By sweeping supply voltage from 0.2 V to 1.2 V, the write margin increases from 34 mV to 320 mV with 9.4 improvement. Utilizing SVT devices in the write paths can generate larger

write margin due to its stronger writeability compared to HVT devices. Although the HVT devices in the 9T cell are relatively weak, they are employed in the entire write paths since compact cell layout for high-density integration and lower leakage is more important. The write failure in the 9T SRAM cell can be compensated by a CAM-assisted write performance boosting

technique.

C. Bitline Leakage Equalization With the Worst Case of Leakage

During read operation, the voltage level of RBL is a function of VDD , device threshold voltage and leakage current, etc. At a specific VDD and a bitline length, the RBL level is highly affected by the amount of leakage current. Maximum bitline leakage occurs when the data in the unselected cells is all logic 0. Similarly, minimum leakage current appears if the data pattern

in the un-accessed cells is all logic 1. Conventionally, a successful bitline sensing requires RBL for data 0 to discharge much faster than that for data 1. However, when variation in bitline

leakage becomes comparable to cell read current, reliable detection of data 0 and 1 is difficult due to the small margin in the current to be sensed. It is shown that, in the worst case, the bitline level of data 0 could be even higher than that of data 1 due to the significant data-dependent bitline leakage particularly at ultra-low voltage.

The conventional bitline sensing problem caused by leakage at ultra-low voltage is illustrated in Fig. 4(a). In read operation,

RWL is enabled and the RBL voltage forms depending on the accessed data. The pull-down strength for sensing 0 should be far higher than that of sensing 1. After that, the simple sense

amplifier (SA) consisting of two stages of invertors senses the voltage of each RBL without trigger timing. As illustrated in

the bottom of Fig. 4(a), this requires the total of the cell current and the minimum leakage current to be far larger than the maximum leakage for successful sensing. As the amount of leakage is comparable to the cell current and the leakage current varies from column to column due to different data pattern, this condition could not be always met.

To address this problem, we propose a bitline leakage equalization technique for single-ended read bitlines. Fig. 4(b) depicts the concept of the proposed bitline equalization technique utilizing the proposed 9T cell. In unselected cells, leakage current flows to GND through the device which is controlled by node QB when the data stored is logic 0. Likewise, when the data is logic 1, leakage current flows to RWL (=GND) through the device controlled by Q. Accordingly, one of two devices connected to Q and QB (M7 and M8 in Fig. 1) is always turned on and the read access device (M9 in Fig. 1) is off. Consequently, two leakage paths have the same strength regardless of the stored data and the constant bitline leakage is formed. In this way, the pull down current for sensing 0 is always larger than that for sensing 1 . This ensures

that the RBL level for data 0 is always lower than that for data 1 and irrespective of the magnitude of . Thus, positive sensing margin could be always provided. Sample simulated

RBL waveforms (Fig. 5) show a drastically improved RBL swing in the 9T SRAM at V whereas the conventional 8T column (HVT(W)-LVT(R)) generates a negligible RBL swing. The proposed scheme improves the RBL swing by

4.6 at 0.2 V, 27 , and 256 cells per bitline. Simultaneously, it also provides a wider sensing timing window. Note that the sensing timing window is defined as the time difference between the RBL of 0 and that of 1 measured when they cross . Since the trip point of our sense amplifier is , we used it as a reference level. With a frequency of 50 KHz, a sensing

timing window of 1.5 s is achieved by the leakage equalization technique whereas nearly no sensing timing window is obtained

in the 8T bitline. The RBL behavior of the 10T SRAM is also captured in Fig. 5. Apparently, the RBLs couldn't fully discharge at this frequency and they are too close to differentiate for sensing.

Variations of cell current and leakage current make RBL swing change and cause sensing problem. Fig. 6 depicts the distribution of RBL swing of the proposed 9T SRAM with 3
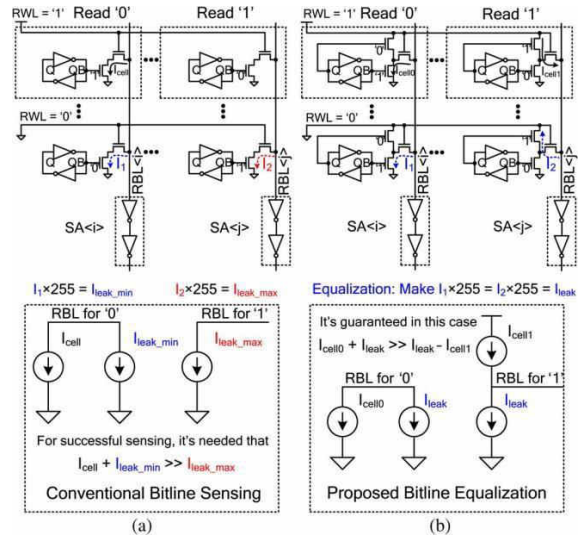

Fig. 4. (a) Conventional bitline sensing in the 8T SRAM [9] and (b) concept of proposed 9T bitline sensing improvement by bitline leakage equalization technique.

local variation at the minimum operating voltage. With a mean value of 53 mV, the RBL swing distribution from 10k-point Monte Carlo runs exhibits a longer right tail. Fig. 7 presents the simulated swing-to- ratio of the proposed 9T SRAM and the 8T SRAM at different temperatures and maximum numbers of cells per RBL (RBL lengths). In order to compare different bitcell topologies in terms of RBL length, we assume nominal process parameter values. In reality, accounting for within-die parametric variations, the effective number of cells per RBL degrades. The proposed 9T SRAM bitline can attach more cells due to the larger RBL swing as verified in Fig. 7. In the 8T SRAM bitline, only 512 cells can be attached for a sensible RBL swing, which at least should be positive. But up to 1024 cells can be attached to the 9T bitline at 0.3 V and 80 . The 8T bitline with 1024 cells generates a negative bitline swing at 80.
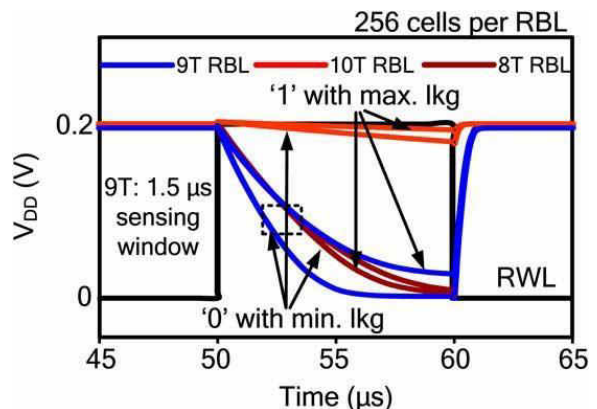
Fig. 5. Improved RBL swing and sensing window of 9T bitline at 0.2 V and fclk=50 KHz with the worst case of leakage.

III. PROPOSED ENERGY EFFICIENT IMPROVEMENT TECHNIQUE

A. *Limitation of MTCMOS on SRAM Energy Efficiency*

At a given SRAM structure, energy efficiency can be optimized by minimizing leakage and maximizing performance. To realize it, the 9T SRAM cell consists of HVT devices in the 6T part and LVT devices in the read port. However, as explained in, this is not the best option in terms of energy efficiency, which is primarily

due to the write performance degradation.

assuming 50% duty cycle.

In the case that is determined by , using HVT in the 6T part reduces and improves the energy. In the other cases, when is determined by , the reduction in and the increase in have to be carefully revisited.

Fig. 8 illustrates a write operation with data flipping. The write operation is divided into two stages, data flipping and data full development. After the data flipping, additional delay is required for the internal nodes, e.g., QB, to be fully developed. In skew conditions (e.g., MTCMOS cell, skew process corners),

QB could move to high voltage very slowly. In this work, the delay till node crossing is defined as the data flipping delay. The delay till data full development (i.e., 90% of V) is defined as

the data full development delay. The latter is more proper to measure the real completion of a write operation. Fig. 8 plots the difference between the data full development delay and the data flipping delay. It is clearly demonstrated that the delay difference between data flipping and full development sharply expands at ultra-low voltage operation.

In an SRAM circuit, the active clock duration is decided by the larger value between the write delay and the read delay. As supply voltage decreases, the write delay with HVT devices degrades faster than the read delay with LVT devices, eventually exceeding the read delay. In this scenario, the overall performance

is limited by the slower write operation. To improve it, the flipping delay instead of the full development delay, can be adopted as write delay when no read-after-write operation is assumed. Fig. 9 shows the issue of the read-after-write operation. After the data flipping at Q and QB, QB rises slowly. When RWL is enabled, Q and QB have not been fully developed yet and the read operation could fail. The write data

could be accessed only after additional clock cycles for full development.

Consequently, the excessively degraded full development

delay nullifies the energy efficiency by prolonging even if significant leakage reduction is achieved with HVT devices in the 6T part. Fig. 10 depicts the read and the write delay of the

9T SRAM array at TT and FNSP corners, respectively. When the supply voltage is lowered below 0.6 V, the full development delay and the flipping delay deteriorate faster than the read

delay [Fig. 10(a)]. The former delay is 6.12 of the latter delayat FNSP corner and V, as shown in Fig. 10(b). The read delay is larger than the flipping delay when is between 0.46 V 0.64 V. In this simulated voltage range, data flipping is

definite to occur within the read delay. Therefore, energy improvement can be obtained if the read-after-write issue could be eliminated by utilizing a faster delay (e.g., read delay/data flipping delay) as .

To address the above issue and enhance energy efficiency, we propose a CAM-assisted circuit for boosting write

performance as well as compensating write failure.

B. *Proposed CAM-assisted Write Performance Boosting Technique*

The slow-write-fast-read problem can be addressed in the architecture level.A completion signal is asserted to alert the CPU when the write operation is finished, otherwise the CPU

stalls for 2 3 cycle during write. Traditional bypass circuit implemented in SRAM utilizing registers to cache input data and its location could also boost performance in the read-after-write case. However, firstly, the register cell can easily cost more than 16 transistors if a main stream DFF style is adopted for ultra-low voltage operation. Secondly, large number of dependent

MUXs and comparators are needed and could not be multiplexed. Therefore, it is beneficial to make the storage circuit, MUXs and comparators implemented with fewer transistors and in an area-efficient array-based way. In this section, we explain a circuit technique that can enhance write performance with this

advantage. Fig. 11 illustrates the operation of the proposed CAM-assisted technique. The SRAM comprises two main paths, an SRAM path and a CAM path. The SRAM path consists of a 16 kb 9T SRAM array (main SRAM array), decoders and data IOs. The CAM path is composed of a tiny 48 b CAM array for storing addresses, a ring counter as an address pointer, an encoder, and a miniature SRAM

array for storing write data. The CAM array (Addr.) and the SRAM array (Data) are implemented with LVT devices for faster read, write, and parallel search to conceal the slow full data development in the main SRAM array. The primary role of the CAM is to store most

recent write addresses and data for possible subsequent read access till the data written into the main SRAM array is fully developed. During write operation [Fig. 11 (left)], data is written into

the main SRAM array (through the SRAM write path) and the miniature SRAM array (through the CAM write path). The write address is stored in the CAM array. The write address and the

data in the CAM can be accessed in the succeeding cycles since the proposed CAM is implemented with LVT devices. During read operation [Fig. 11(right)], the main SRAM array is accessed for normal read operation, and the CAM array is simultaneously searched using the read address as search data. If the read address is not found in the CAM array, the cells

that are written in the preceding cycles couldn't be accessed. Thus, the selection signal from the encoder will select the read data from the main SRAM array as the final data through MUX. If an address match occurs by a subsequent read-after-write operation, the encoder enables a wordline signal corresponding to the matched address. The wordline activates reading data from the SRAM array and later the data is sent to MUX. Finally, using the selection signal from the encoder MUX will select the data from the proposed CAM as the final data. In this case, the read data from the main SRAM array cannot be used as the final data since the data written in the previous cycle has not been fully developed

due to the slow development speed of the latches using HVT devices. Therefore, the read data from the CAM should be selected as the final read data. Through this, the write performance is determined by the read operation or the data flipping delay, not by the slower full development delay. As a result,

instant read-after-write operation for the same address is executable without slowing down the clock frequency for providing full data development in the main SRAM array.

Fig. 12 compares the delays of four different operations (i.e., SRAM read, SRAM write, CAM read, and CAM write) to demonstrate the performance advantage of the proposed

scheme. The delay of SRAM write is calculated by the full development time. As shown in Fig. 12, the delay of SRAM write is the largest whereas that of CAM write is the smallest. Since

the CAM-assisted technique hides the slow SRAM write, the overall performance is improved from SRAM write to SRAM read. The performance improvement of 47.5% is achieved from simulation.

The schematic and the searching operation of the 10T CAM cell employed in this work adopts from.The CAM cellcomprises a 6T SRAM part and search logic circuits. Before search operation, the match line (ML) is precharged to .
A search operation starts by loading search data into the search lines. If the search data is different from the stored data, one of the search
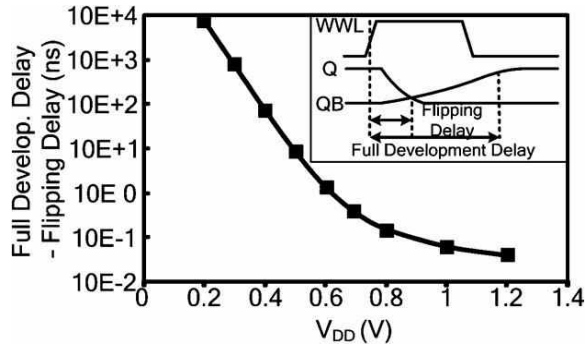

Fig. 8. Definition of data flipping delay and data full development delay. Difference of the data flipping delay and the full development delay substantially increases with scaling VDD
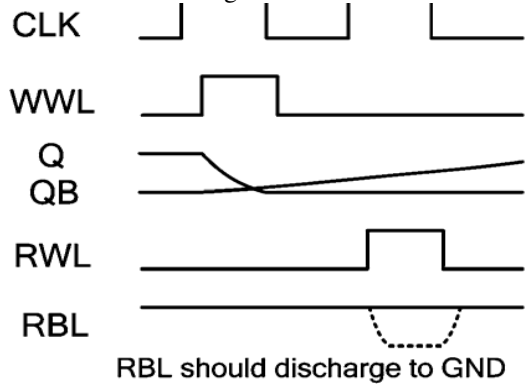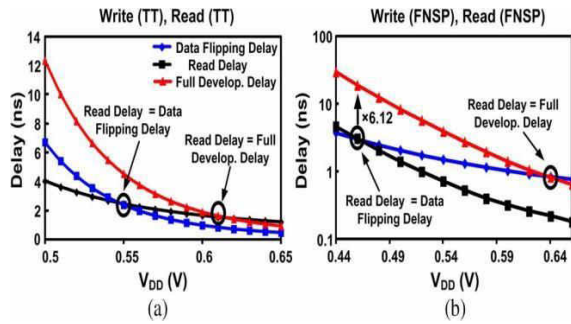

Fig. 9. Read failure due to data non-full development in SRAM cell nodes.

Fig. 10. (a) Read and write delays against scaling at TT corner. (b) Read and write delays against scaling at FNSP corner.
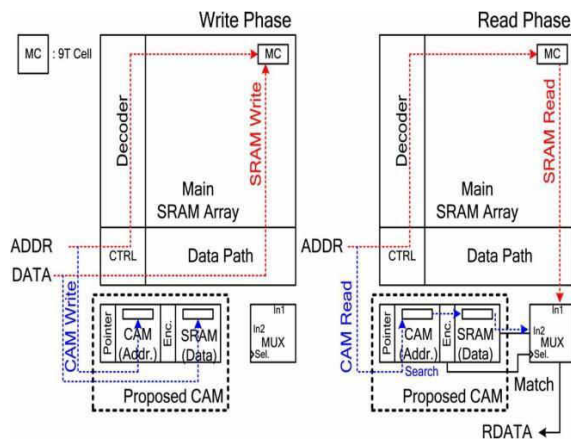

Fig. 11. Data paths of write and read operations in the CAM-assisted SRAM circuit.

logic circuits would discharge ML to GND. Contrarily, if the search data is identical to the stored data, ML remains at a high voltage. The circuit diagram of the CAM-assisted circuit is described in Fig. 13. Conventionally, input of a CAM is data and output is a hit address. In this work, input is a read address and output is data. The CAM array is comprised of 4 rows (i.e., storing 4 most recent write addresses) and 12 columns (i.e., 12-bit address). The number of rows is mainly determined by the ratio of the data full development delay and the flipping delay. A ring counter is utilized to act as a pointer for the CAM array. When a write operation is asserted, the pointer enables one row, writing the input address into the CAM array and the data into the miniature SRAM array. When a SRAM read operation is enabled, the address is loaded into the search lines (SL and SLB) of the CAM array.
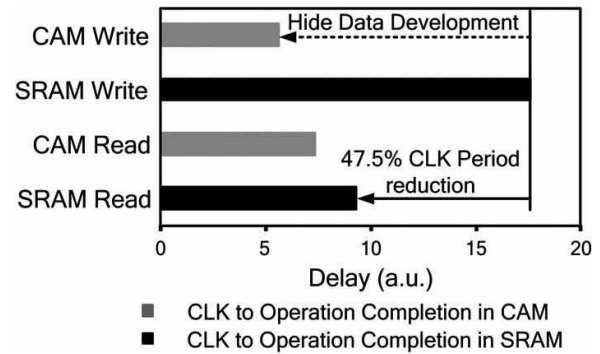

Fig. 12. Delay of four different operations of SRAM and CAM circuits.

If the address is found from the CAM array, the corresponding ML(s) will be enabled. Otherwise, no ML is enabled and the search operation finishes. If multiple MLs are enabled, the encoder activates only one read wordline (CAM_RWL) corresponding to the most recent write operation. The activated wordline enables reading data through read bitlines (CAM_RBL) and sending the read data to MUX (Fig. 11). The number of rows in the CAM array. can be estimated by the following equation if 50% clock duty cycle is assumed If in Fig. 14 is greater than 2, read operation is likely to fail in the subsequent read operation (50% duty cycle), which is addressed by the proposed CAM-assisted technique. To cover a case at FNSP corner [Fig. 10(b)], should be at least , which is 3. In this work, we implemented 4 rows to provide a redundancy in for real application.
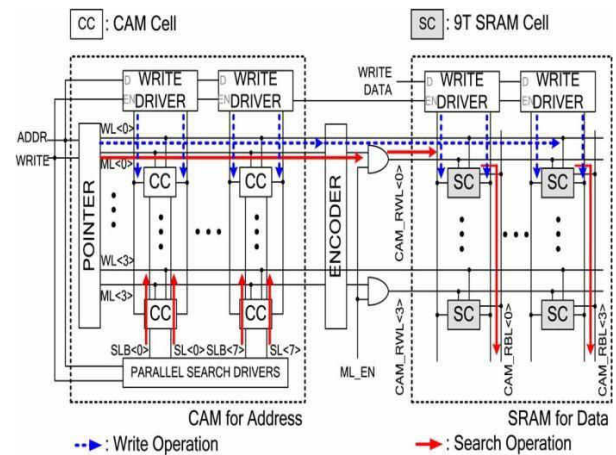

Fig. 13. Circuit diagram of CAM array, search logic, and miniature SRAM array.
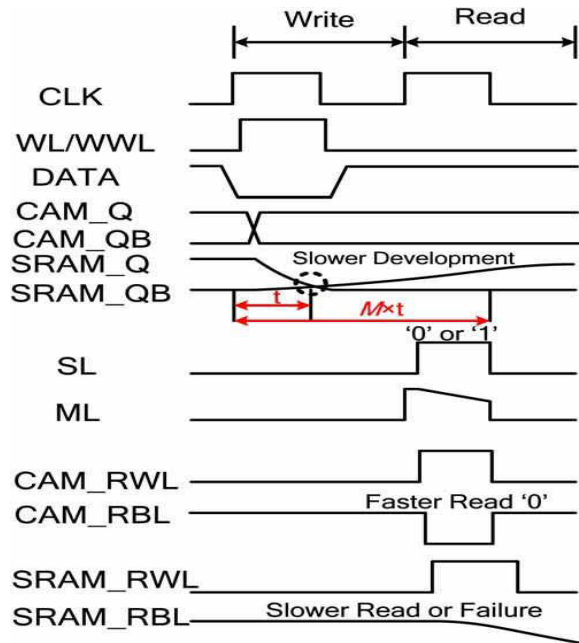
Fig. 14. Timing diagram of SRAM array and CAM circuit during succession of write and read operations.

The timing diagram of the proposed CAM-assisted SRAM is illustrated in Fig. 14. The data in the tiny SRAM (CAM_Q/QB) develops much faster than that in the main SRAM array (SRAM_Q/QB). In the subsequent read operation, input address in the search lines (SLs) keeps the corresponding ML high, and accordingly quickly generates CAM_RWL and CAM_RBL due to the LVT devices and the small load. The other read path through RWL generates SRAM_RBL with a larger delay and, in the worst case, it generates a read failure.
Fig. 15 manifests that the full development delay of the CAM is always smaller compared to the main SRAM array at all corners. Simultaneously, the full development delay of the CAM is also shorter than the read delay of the main SRAM array, making the read paths critical.

IV.   TEST CHIP IMPLEMENTATION AND MEASUREMENT
The main SRAM array is organized with 256 words4 bits 4, which occupies an area of 169 195 (including power rails in rows and columns). It is divided into 4 sub-blocks and each sub-block is composed of 16 columns, sharing one IO. The CAM array is configured with 4 rows and each row has 12 CAM cells for storing addresses and 4 SRAM cells (LVT) for storing write data. The proposed CAM circuit occupies 1061 (not including interconnections), which is at least 60% smaller than the DFF-based design in our estimation.
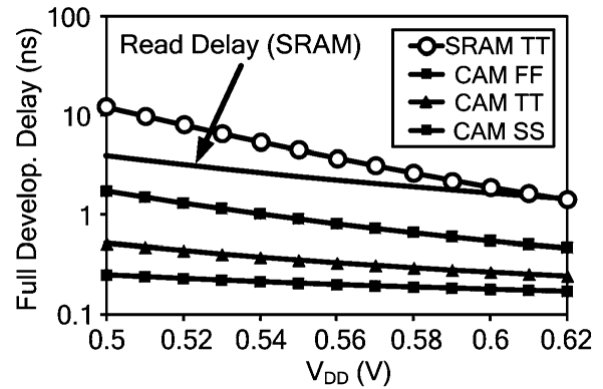

Fig. 15. Faster write completion in CAM array than SRAM array at different corners.

It causes an overhead approximately as 3% of the SRAM array area. The overhead will be less at a higher SRAM
array density since the number of rows is mostly determined by a single cell. The energy dissipation by the proposed CAM circuit occupies a very small portion of the overall consumption. Simulation shows that the CAM energy per read with search operation is 59 fJ at 0.4 V with frequency of 1 MHz. To be more flexible, data from CAM_RBL and SRAM_RBL can bypass the MUX for separate measurement.
A 16 kb SRAM test chip is fabricated in a commercial 65 nm CMOS technology with a nominal of 1.2 V. Fig. 16(a) shows the experimental results of the leakage current.
At 27 , the leakage current of the test chip changes from 139. 3 A (1.2 V) to 1.4 A (0.1 V). When temperature goes to 100 , it becomes 305 A and 4.5 A, respectively. Power of read and write operation is measured at the maximum
operating frequency [Fig. 16(b)]. The read power is larger than the write power due to the precharging and discharging current in the read bitlines. The average power is measured in the supply range of interest, assuming equal probability of performing read and write operations. It changes from 146 W at 0.6 V to 4.12 W at 0.32 V. Fig. 17(a) verifies that the CAM circuit can provide a shorter read access time by 25.8% at 0.6 V and 7% at 0.38 V compared to the SRAM without the CAM.
Below 0.38 V, read from CAM takes more time due to slow search operation at ultra-low voltage. The operating frequency of the CAM-assisted SRAM is depicted in Fig. 17(b). Around
the critical voltage of 0.6 V, the CAM circuit speeds up clock frequency of the main SRAM to 40 MHz. The maximum operating frequency at is boosted to 5 MHz. The SRAM performance is therefore improved

by 42.6% and 66.7% at 0.6 V and 0.4 V, respectively. The plot of energy per
operation is shown in Fig. 18. The SRAM consumes an energy per operation of 3.47 pJ at . Thanks to the CAM-assisted circuit, a minimum energy per operation of 2.07 pJ is achieved and the energy efficiency is consequently improved by 40.3%. Averagely, the energy efficiency in the supply range of 0.38 V to 0.6 V is enhanced by 29.4%. The test cells are fully functional down to 0.26 V with the maximum operating
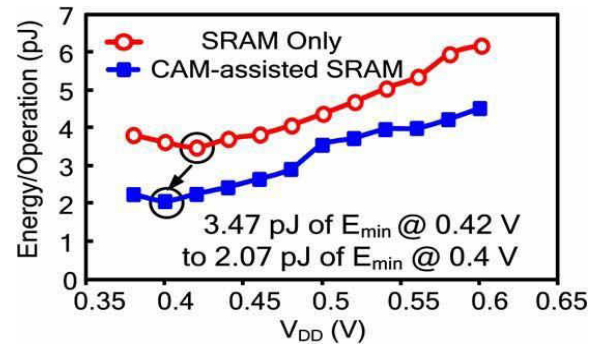

Fig. 18. Measured energy of SRAM only and the CAM-assisted SRAM.
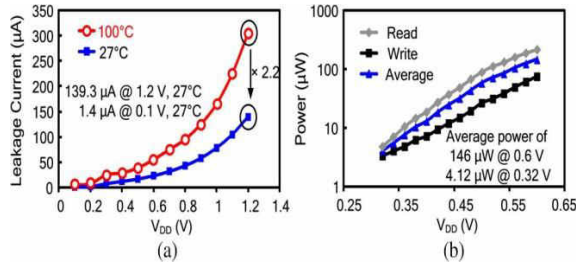

Fig. 16. Measured (a) leakage current of the test chip and (b) write, read, and average power at maximum operating frequency.
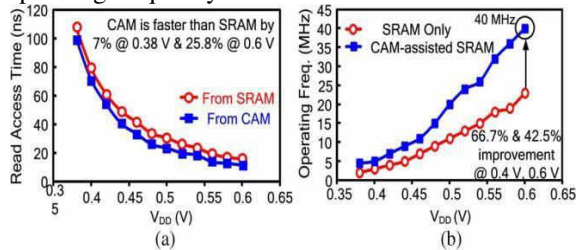

Fig. 17. Measured (a) read access time and (b) improved operating frequency of the CAM-assisted SRAM.

frequency of 100 kHz (27). The read access time of the SRAM is measured as 0.85 s at 0.26 V. The CAM circuit achieves an average improvement of 18.7% in the read access time between 0.38 V 0.6 V. It lowers the minimum read voltage further from 0.26 V to 0.23 V. The test chip micrograph with waveform capture is shown in Fig. 19. Table I compares the test chip with various ultra-low voltage SRAM circuits. Among the SRAMs, this work achieves the lowest minimum energy when it is normalized with respect to density.
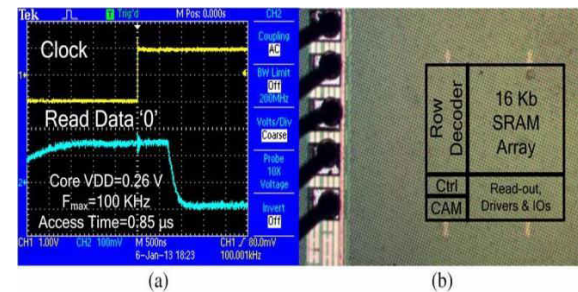

Fig. 19. (a) Readout waveforms capture at 0.26 V. (b) Die microphotograph.

V. CONCLUSIONS

Leakage and energy efficiency are primary concerns for ultra-low voltage SRAM design. This paper presents several circuit techniques to implement an energy efficient SRAM with reliable read operation under ultra-low voltage. The proposed 9T SRAM cell with equalized bitline leakage fosters SRAM read operation at ultra-low voltage, achieving read access time of 79 ns at 0.4 V and 0.85 s at 0.26 V, respectively. To further reduce the static energy, MTCMOS technology is utilized to reduce the leakage in the SRAM array. While HVT devices in the 6T part reduce leakage, they degrade write performance significantly at low voltage. This nullifies the energy efficiency improvement in the near- or the sub-threshold region. To tackle this issue, we proposed a CAM-assisted write performance boosting circuit to speed up clock frequency. The test chip shows an average performance improvement of 29.4% with the aid of the proposed circuit technique. Consequently, the energy efficiency is improved by 40.3% with the minimum energy
per operation of 2.07 pJ at 0.4 V. The measurement results prove that the proposed techniques are good circuit solutions for ultra-low voltage and energy efficient applications.

TABLE
DESIGN METRIC COMPARISON WITH VARIOUS ULTRA-LOW VOLTAGE SRAMS

	A-SSCC 2012 [14]	JSSC 2013 [15]	JSSC 2013 [16]	This Work
Technology	65 nm	65 nm	65 nm	65 nm
Density	128 kb	2 kb	32 kb	16 kb
Transistor count	8T	9T	7T	9T
Cell size	N.A.	$1.24 \times 2.31 \ \mu m^2$	N.A.	$2.63 \times 0.72 \ \mu m^2$
V_{DDmin}	0.37 V	0.28 V	0.26 V	0.26 V
Access time	N.A.	4.55 µs (0.3 V)	0.55 µs (0.26 V)	0.85 µs (0.26 V)
Leakage current	N.A.	0.05 µA (0.4 V)	N.A.	1.4 µA (0.1 V)
Min. energy (E_{min})	21.2 pJ	0.57 pJ	5.6 pJ	2.07 pJ
Normalized E_{min}	162 aJ/b	278 aJ/b	171 aJ/b	126 aJ/b

REFERENCES
[1] H. Kim *et al.*, ‒A reconfigurable SIMT processor for mobile ray tracing with contention reduction in shared memory,‖,Apr. 2013.

[2] M. Ghaed *et al.*, ‒Circuits for a cubic—millimeter energy—autonomous wireless intraocular pressure monitor,‖ Dec. 2013.

[3] M. Tu *et al.*, ‒A single-ended disturb-free 9T subthreshold SRAM with cross-point data-aware write word-line structure, negative bit-line, adaptive read operation timing tracing,‖ Jun. 2012.

[4] A. Teman *et al.*, ‒A 250 mV 8 kb 40 nm ultra-low power 9T supply feedback SRAM (SF-SRAM),‖ Nov. 2011.

[5] T. Kim *et al.*, ‒A voltage scalable 0.26 V, 64 kb 8T SRAM with lowering techniques and deep sleep mode,‖ Jun. 2009.

[6] A. Alvandpour *et al.*, ‒Bitline leakage equalization for sub-100 nm caches,‖ 2003.

[7] D. Jeon *et al.*, ‒A super-pipelined energy efficiency subthreshold 240 /s FFT core in 65 nm CMOS,‖Jan. 2012.

[8] R. Abdallah and N. Shanbhag, ‒A 14.5 fJ/cycle/k-Gate, 0.33 V ECG processor in 45 nm CMOS using statistical error compensation,‖ Sep. 2012.