

A 140-mV Variation-Tolerant Deep Sub-Threshold SRAM in 65-nm CMOS

Khawar Sarfraz, *Member, IEEE*, Jin He, *Member, IEEE*, and Mansun Chan, *Fellow, IEEE*

Abstract—This paper presents a sub-threshold SRAM, which eliminates bitline (BL) leakage-induced read failures. The proposed architecture clamps the current ratio between differential BLs to a fixed value, thus permitting reliable ultra-low-voltage read-out. A de-multiplexed wordline interleaving scheme is presented to compensate for bitcell area overhead. The interleaving technique achieves 9% reduction in decoder area and 50% reduction in clock load within the decoder. A sense amplifier circuit with reduced sensitivity to process variations is proposed to further enhance the reliability of the differential read-out. Measurement results from a 1-kb SRAM, fabricated in an industrial 65-nm low-power CMOS process, show 13.1-kHz operation at 140 mV, with active read and leakage power figures of 30.5 and 28.1 nW, respectively.

Index Terms—Differential bitcell, low voltage, memory, ultra-low power.

I. INTRODUCTION

IMPLANTABLE devices and wireless sensor nodes are severely energy-constrained platforms that rely on extreme voltage scaling for prolonged battery life. Reliable operation of digital circuits for these platforms has been shown in the near- and sub-threshold regions with ultra-low power consumption [1]. Achieving reliable sub-threshold SRAM functionality for these platforms is, however, challenging primarily due to the large number of leakage current paths in read bitlines (BLs), which leads to diminishing noise margins, and consequently to read failures [2], [3]. Column data randomization [4] enhances noise immunity of BLs at the cost of considerable area and power overhead. The 216-mV SRAM [3] uses 21-transistor (21T) bitcell, single BL evaluation transistor, and contention-free keeper. The 180-nm SRAM [5] uses 14T bitcell with complementary-stacked read port for robust reads and low standby power. BLs designed with PMOS-embedded 4T read port [2] experience an inflow of leakage currents from non-accessed bitcells, which may prevent a valid evaluation. The 6T bitcell [6] is prone to a flip due to direct read access mechanism. Moreover, a BL misevaluation may occur if the worst-case read “0” and read “1” delays become comparable. Similarly, a differential read [7] may

Manuscript received January 4, 2017; revised April 10, 2017; accepted May 9, 2017. Date of publication June 20, 2017; date of current version July 20, 2017. This work was supported by RGC Hong Kong under Grant N_HKUST605_12. This paper was approved by Associate Editor Vivek De. (*Corresponding author: Khawar Sarfraz*.)

K. Sarfraz and M. Chan are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: ksfraze@connect.ust.hk; mchan@ust.hk).

J. He is with the Shenzhen SOC Key Laboratory, Peking University, Shenzhen 518057, China (e-mail: frankhe@pku.edu.cn).

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

Digital Object Identifier 10.1109/JSSC.2017.2707392

fail when cumulative leakage currents on non-evaluated BL become comparable to read current on the evaluated BL.

In this paper, BL leakage-induced read failures are eliminated by fixing the current ratio between differential BLs at a constant value, which is independent of column height, the size of evaluation transistors, and column data configuration as described in Section II. A de-multiplexed wordline interleaving scheme, discussed in Section III, reduces clock load and compensates for bitcell area overhead. The voltage and time sensing margins of the SRAM are further enhanced with a variation-tolerant sense amplifier (SA), as explained in Section IV. Test chip measurements are discussed in Section V, and the conclusion is provided in Section VI.

II. ELIMINATING BITLINE LEAKAGE-INDUCED READ FAILURES

The proposed bitcell, shown in Fig. 1, uses the latch in [8] except that write BL (WBL) is tri-stated in standby. Leakage-biased WBLs achieve up to 14.3% reduction in column leakage currents under worst-case data pattern, i.e., when 50% bitcells store “1.” The WBL driver, shown in Fig. 1, is shared by two adjacent columns. The least significant write address bit WrAddr[LSB] selects between two interleaved words. The active-high global write enable input is de-asserted in the absence of writes to tri-state WBL(A) and WBL(B).

BL evaluation transistors, M4 and M8, are cut off when the read wordline (RWL) is inactive (high). If “1” is stored on Q, M4 is turned OFF by the voltage on RWL since M2–M3 are turned ON. M8 is, however, cut off by the turned-ON M5. During reads, RWL is asserted low to discharge node X. The pre-discharged BL is consequently charged by the turned-ON M4, while complementary bitline (BLB) is maintained at 0 V. M4 compensates for the additional delay that is incurred in discharging node X since it does not suffer from the stack effect. M4 and M8 contribute 10% to bitcell leakage power, while M1–M3 and M5–M7 do not produce source-drain leakage currents. M4 and M8 are isolated from other bitcell transistors by a separate N-well that is shared with the adjacent column, as shown by the layout in Fig. 1. The use of a separate N-well permits optional speed gain with the application of forward body bias (FBB) exclusively to M4 and M8.

The proposed data column is shown in Fig. 2(a). Assuming BL is charged in a read cycle, the current flowing into BL is $I_R + (n - 1) \cdot I_L$, where n is the column height, I_R is the read current, and I_L is the source-drain leakage current of M4 and M8. Alternatively, the current flowing into BLB is $(n) \cdot I_L$. The current difference between BL and BLB is thus fixed at $I_R - I_L$, which is independent of column height,

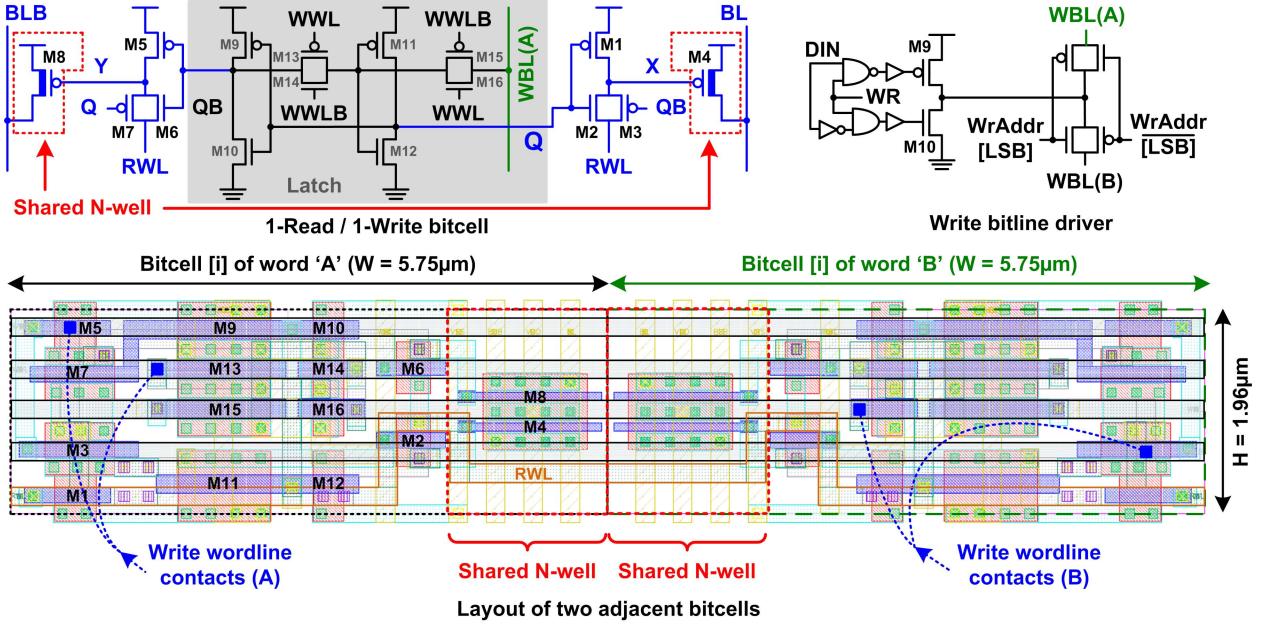


Fig. 1. Proposed 1-read/1-write bitcell, WBL driver that supports leakage-biased WBLs, and the layout of two-bitcell repeatable unit. M4 and M8 are high threshold voltage (high V_t) devices. All other devices are standard V_t .

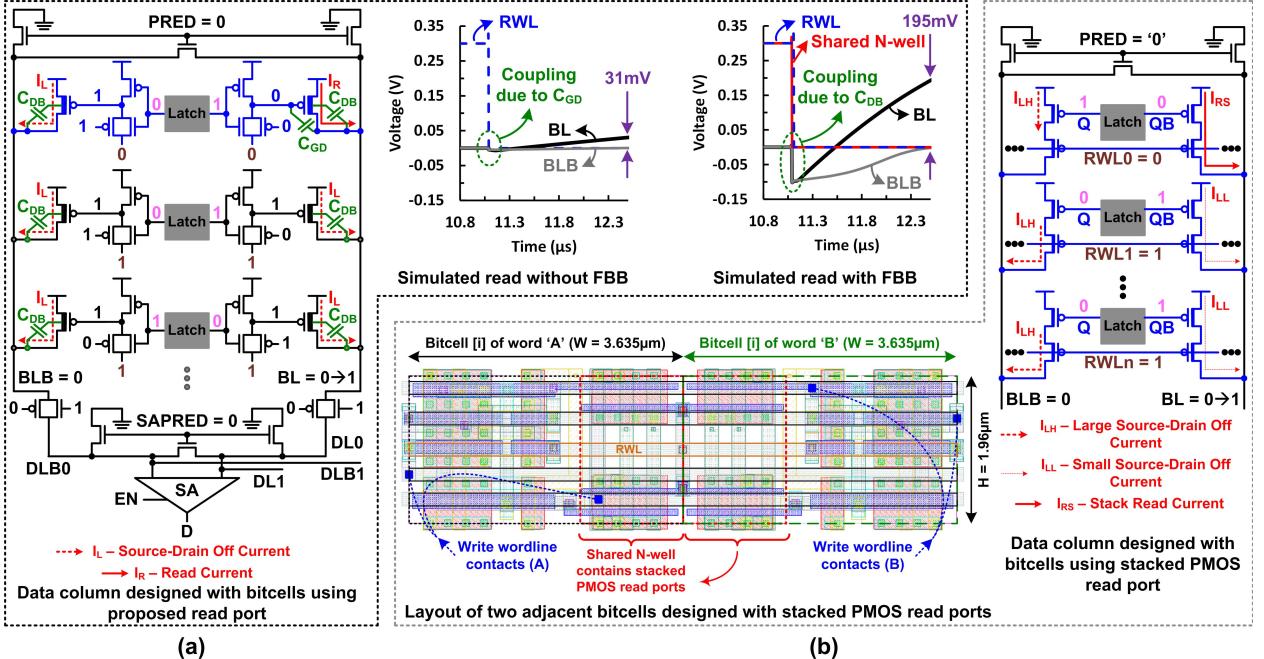


Fig. 2. (a) Illustration of BL leakage currents in a data column designed with the proposed bitcell. Transient read waveforms illustrate the formation of BL differential voltage with and without the introduction of FBB. Typical Corner, $V_{DD} = 0.3$ V and $T = 25$ °C. (b) Illustration of BL leakage currents in a data column designed with bitcells using stacked PMOS differential read port. Layout of two-bitcell repeatable unit designed with stacked PMOS differential read port.

the strength of M4 and M8, and column data configuration. Fixing the current difference between BL and BLB is key to eliminating leakage-induced read failures. The proposed read port achieves this objective by isolating the gates of evaluation transistors from bitcell storage nodes and by ensuring that source-drain leakage currents of M4 and M8 are equal for non-accessed column bitcells. The application of FBB to M4 and M8 enhances the read frequency by a factor of 4

at 0.3 V, as illustrated by read waveforms in Fig. 2(a). Separate buffers switch the shared N-wells from supply voltage level to 0 V in sync with the falling edge of RWL.

A data column designed with stacked PMOS read ports is shown in Fig. 2(b) for comparison. At low supply voltages and under worst-case conditions and data configuration, the current flowing into BLB ($(n) \cdot I_{LH}$) may become comparable to the current flowing into BL ($I_{RS} + (n - 1) \cdot I_{LL}$) for a certain n .

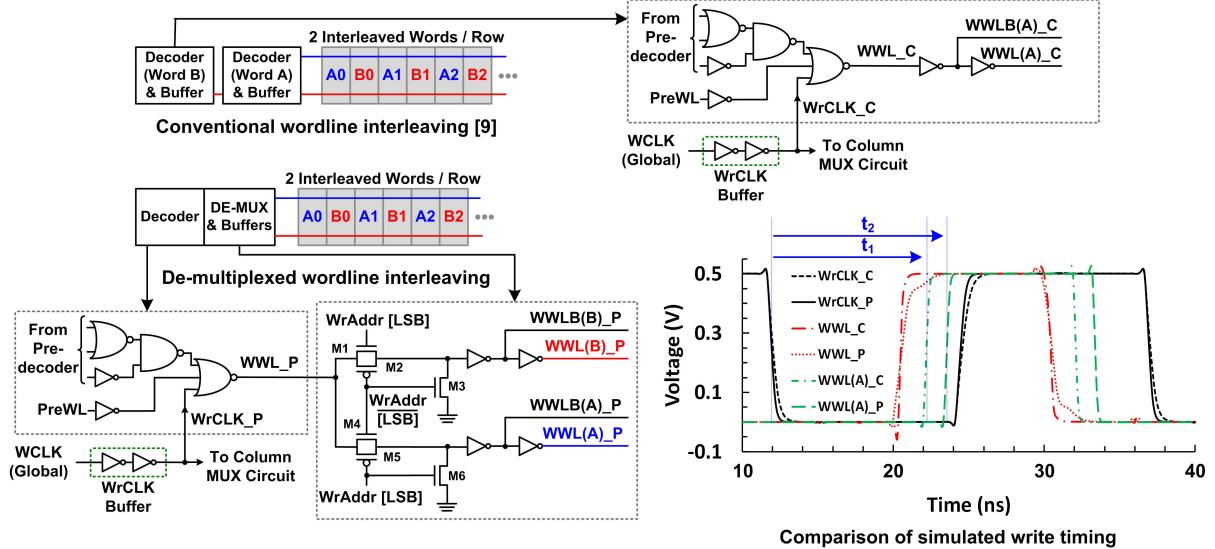


Fig. 3. Conventional and de-multiplexed write wordline interleaving scheme, with a comparison of simulated write timing under typical process corner, $V_{DD} = 0.5$ V and $T = 25$ °C.

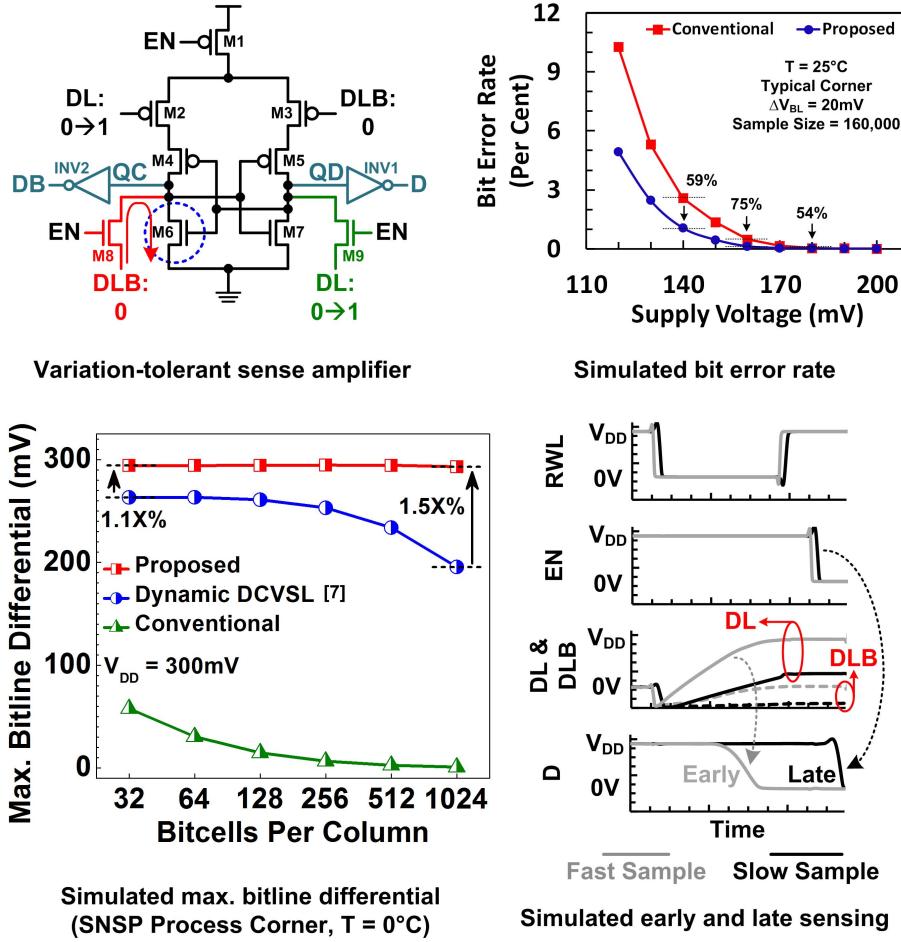


Fig. 4. Schematic of variation-tolerant SA and simulations demonstrating improved performance.

BL and BLB could therefore experience a similar rate of charge, which may result in failure of a differential read. Compared to the proposed bitcell, stacked PMOS read ports lead to 36.8% smaller bitcell, as shown in Fig. 2(b), and 14.2% lower bitcell leakage currents.

III. DE-MULTIPLEXED WORDLINE INTERLEAVING

The conventional [9] and proposed interleaving schemes are shown in Fig. 3. Instead of two adjacent decoders with individual buffers, the proposed scheme uses one decoder

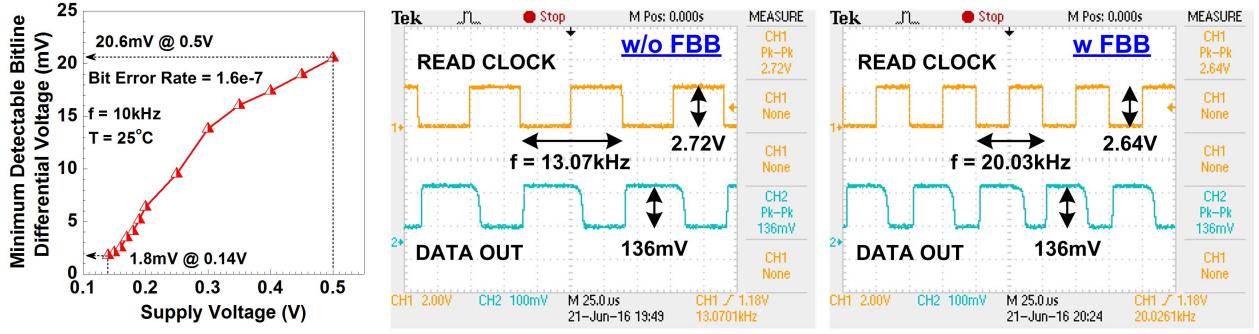


Fig. 5. SA measurements and oscilloscope read traces at a supply voltage of 140 mV and $T = 25^{\circ}\text{C}$.

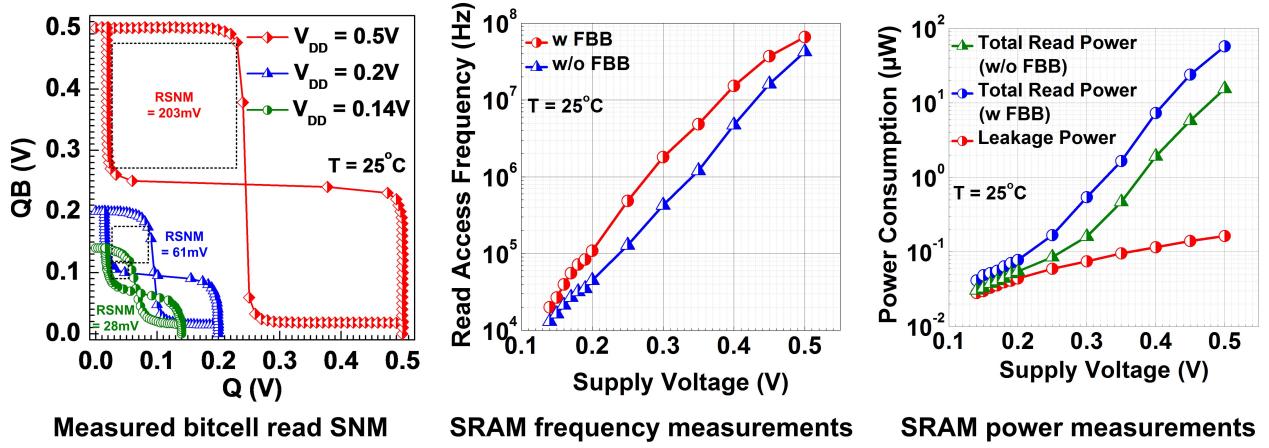


Fig. 6. Measured butterfly curves and SRAM frequency and power measurements.

and de-multiplexer followed by individual buffers to route decoded WWL_P to the selected interleaved word. When WrAddr[LSB] is “0,” WWL_P is routed to WWL(A)_P since M4 and M5 are tuned ON. Alternatively, WWL(B)_P is maintained at 0 V by the turned-ON M3. The proposed design permits 50% reduction in write clock (WrCLK_P) load within the decoder and achieves a 9% reduction in decoder area for two interleaved words. In the write timing diagram of Fig. 3, the rise/fall time for WrCLK_P is reduced by 36% as compared to that for WrCLK_C due to smaller gate capacitance and wiring load. The rise/fall time for WWL_P is, however, 4× larger as compared to WWL_C due to additional loading of de-multiplexer logic. WWL(A)_P is consequently delayed ($t_2 - t_1$) by 5.4% of clock period, which does not impact the read-limited SRAM frequency.

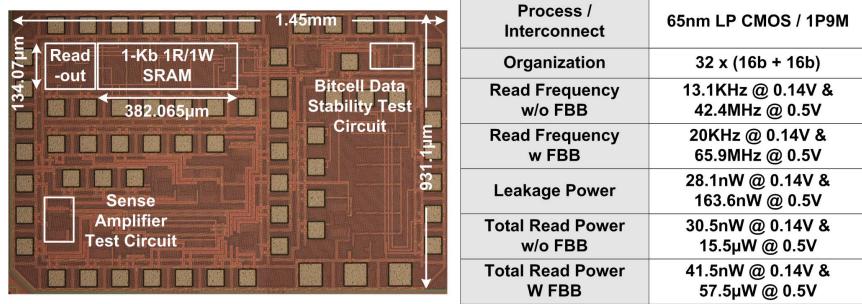
IV. VARIATION-TOLERANT SENSE AMPLIFIER

The variation-tolerant SA is shown in Fig. 4. Contrary to conventional design [10], the source terminals of M8 and M9 are disconnected from ground and instead connected to complementary dataline (DLB) and dataline (DL), respectively, a change that results in an area penalty of 16.8%. Nodes QC and QD are thus held at the developed DL differential prior to assertion of the active-low enable (EN) signal. The voltage on DL is transferred to the gates of M4, M6, and M2. Similarly, voltage on DLB is transferred to the gates of

M5, M7, and M3. Assuming DL is charged in a read cycle, QD is preset to a positive voltage before EN is asserted low. This positive voltage on QD is closer to the desired final voltage on QD, i.e., after EN is asserted low. The bit error rate (BER) of proposed SA is thus 59% lower at 140 mV, as shown in Fig. 4. The BER is estimated for a total of 16 SAs using a target yield of 99.99% [11]. Furthermore, the rising voltage on QD gradually turns ON M6. DLB, which is slightly charged by leakage currents, is consequently discharged back to ground via M8 and M6, thereby enhancing the voltage sensing margin. Due to lower absolute threshold voltage and increased drive strength under an equal area constraint, M6 in Fig. 4 turns ON earlier compared to the PMOS keeper in [7]. The maximum DL differential is thus enhanced by 1.5× at 300 mV with 1024 bitcells per column compared to [7], as shown in Fig. 4. Since one of the DLs is eventually clamped to a near-zero potential prior to the trigger event, the time margin that is associated with EN is relaxed. Moreover, DL samples that evaluate faster (early) are correctly transferred to D via M9 and INV1 prior to assertion of EN, as shown in Fig. 4. Late (typical) DL samples are correctly sensed when EN is asserted low.

V. CHIP FABRICATION AND MEASUREMENTS

The techniques described in Sections II–IV are implemented in a 1-kb SRAM that is fabricated in a 65-nm low-power



Die photo and chip performance summary

Metric	[4]	[3]	[5]	[2]	[6]	[7]	[12]	[13]	This Work
Process	65nm	130nm	180nm	130nm	130nm	90nm	65nm	65nm	65nm
Bitcell Type / Area	8T / 1.352μm ²	21T / -	14T / 40μm ²	10T / 7.504μm ²	6T / 4.788μm ²	10T / -	16T / -	17T / -	16T / 11.27μm ²
Bitcell Area Overhead	-	-	9.1X wrt 6T Bitcell	-	2.0X wrt 6T Bitcell	1.61X wrt 8T Bitcell	-	-	9.0X wrt 6T Bitcell
V _{MIN} w/o R/W Failures	200mV	216mV	500mV	200mV	210mV	180mV	420mV	350mV	140mV
f _{MAX} @ V _{MIN}	400KHz	28KHz	100KHz	120KHz	21.5KHz	500Hz	10KHz	8KHz	13.1KHz
Array Size	32Kb	6.7Kb	3.4Kb	480Kb	2Kb	32Kb	4Kb	4Kb	1Kb

Comparison with prior work

Fig. 7. Die photograph, chip performance summary, and comparison with prior work.

CMOS test chip. The array is designed with 16 SAs, 32 rows, and 32 columns. Each row has two 16-bit interleaved words. Read BLs and WBLs are 32 rows deep. Standard- V_t logic gates, with long channel length, are sized for symmetrical voltage transfer characteristics using worst-case process corner and Monte Carlo analyses. SA measurements in Fig. 5 are performed using a BER of $1.6e - 7$ and the confidence level of 99%. The SA fails below 140 mV. Traces in Fig. 5 illustrate correct read-out after a known data combination is written to specific addresses at 140 mV. No read bit errors are observed at the stated read frequency values. Measured butterfly curves in Fig. 6 demonstrate bitcell read static noise margin of 28 mV at 140 mV. The measured read access frequency increases with the introduction of FBB, as shown in Fig. 6, following the argument in Section II. The measured read power in Fig. 6 represents power consumed in accessing a 16-bit word and includes the power consumed in bitcell array, peripheral logic, timing block, and 16 SAs. The measured leakage power represents standby power consumption of the entire SRAM. The minimum operating voltage is limited by the PMOS-input SA, and not the read BLs. The die photograph, chip performance summary, and a tabulated comparison with prior work are shown in Fig. 7.

VI. CONCLUSION

A deep sub-threshold SRAM is presented for reliable low-voltage operation. A differential read port is proposed to eliminate BL leakage-induced read failures. A de-multiplexed wordline interleaving scheme reduces clock load within the decoder and compensates for the increase in bitcell area. A variation-tolerant SA is presented for robust differential read-out. Measurements performed on a 65-nm test chip demonstrate correct read operation at 140 mV.

REFERENCES

- [1] M.-E. Hwang, A. Raychowdhury, K. Kim, and K. Roy, "A 85mV 40nW process-tolerant subthreshold 8×8 FIR filter in 130nm technology," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2007, pp. 154–155.
- [2] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A high-density subthreshold SRAM with data-independent bitline leakage and virtual ground replica scheme," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 330–606.
- [3] J. Chen, L. T. Clark, and T.-H. Chen, "An ultra-low-power memory with a subthreshold power supply voltage," *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2344–2353, Oct. 2006.
- [4] A. T. Do *et al.*, "0.2 V 8T SRAM with PVT-aware bitline sensing and column-based data randomization," *IEEE J. Solid-State Circuits*, vol. 51, no. 6, pp. 1487–1498, Jun. 2016.
- [5] S. Hanson *et al.*, "A low-voltage processor for sensing applications with picowatt standby mode," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1145–1155, Apr. 2009.
- [6] B. Zhai, D. Blaauw, D. Sylvester, and S. Hanson, "A sub-200mV 6T SRAM in 0.13 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Mar. 2007, pp. 332–606.
- [7] I. J. Chang, J.-J. Kim, S. P. Park, and K. Roy, "A 32kb 10T subthreshold SRAM array with bit-interleaving and differential read scheme in 90nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2008, pp. 388–622.
- [8] K. Sarfraz and M. Chan, "A 1.2V-to-0.4V 3.2GHz-to-14.3MHz power-efficient 3-port register file in 65-nm CMOS," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 2, pp. 360–372, Feb. 2017.
- [9] H. Fujiwara *et al.*, "A 20nm 0.6V 2.1μW/MHz 128kb SRAM with no half select issue by interleave wordline and hierarchical bitline scheme," in *Symp. VLSI Circuits Dig. Tech. Papers*, 2013, pp. C118–C119.
- [10] B. Liu, J. Cai, J. Yuan, and Y. Hei, "A low-voltage SRAM sense amplifier with offset cancelling using digitized multiple body biasing," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 64, no. 4, pp. 442–446, Apr. 2017.
- [11] A. Singhee and R. A. Rutenbar, "Statistical blockade: Very fast statistical simulation and modeling of rare circuit events and its application to memory design," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 8, pp. 1176–1189, Aug. 2009.
- [12] P. Meinerzhagen *et al.*, "A 500 fW/bit-access 4kb standard-cell based sub-VT memory in 65nm CMOS," in *Proc. Eur. Solid-State Circuits Conf.*, 2012, pp. 321–324.

- [13] O. Andersson, B. Mohammadi, P. Meinerzhagen, A. P. Burg, and J. N. Rodrigues, "Dual-VT 4kb sub-VT memories with <1 pW/bit leakage in 65 nm CMOS," in *Proc. Eur. Solid-State Circuits Conf.*, 2013, pp. 197–200.



Khawar Sarfraz (S'07–M'10) received the B.Sc. degree in electronic engineering from GIK Institute of Engineering Sciences and Technology, Topi, Pakistan, in 2000, the M.Sc. degree in electrical engineering (microelectronics) from Delft University of Technology, Delft, The Netherlands, in 2009, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2016.

He has worked for the electronic hardware design industry for over 8 years and has more than 2 years of academic teaching experience. His current research interests include high-performance and low-power single- and multi-port embedded memories.



Jin He (M'04) received the B.S. degree in electrical engineering from Tianjin University, Tianjin, China, in 1988, and the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1993 and 1999, respectively.

From 2001 to 2005, he was a Research Engineer with EECS, University of California, Berkeley, CA, USA. Since 2005, he has been a Full Professor with Peking University, Beijing, China. In 2006 and 2014, he was a Visiting Scholar with

The Hong Kong University of Science and Technology, Hong Kong. In 2008, he joined Hiroshima University, Hiroshima, Japan, as a Visiting Professor, and he was a Guest Professor at Chinese Academy Science, Beijing, from 2008 to 2011. Since 2010, he has also been a Full Professor with the Peking University Shenzhen Institute, and the Director of Shenzhen System-on-Chip Key Laboratory, PKU-HKUST Shenzhen-Hong Kong Institution, Peking University. He was one of the main contributors of the international standard CMOS model BSIM4.3.0, which is widely used in the international semiconductor industry field. He is involved into a series of the National major state basic research projects, such as 973, 863, NSFC, and a number of SRC projects. He was an Invited Speaker of over 60 invited talks to international conferences, societies, universities, and industry. He has authored more than 200 journal and 200 conference papers and edited five books. His current research interests include process technology, transport physics, modeling and simulation of nano-scale devices in future-generation information processing/storage.



Mansun Chan (S'92–M'95–SM'01–F'13) received the B.S. (Highest hons.) degree in electrical engineering and the B.S. (Highest hons.) degree in computer sciences from the University of California, San Diego, CA, USA, in 1990 and 1991, respectively, and the M.S. and Ph.D. degrees from the University of California, Berkeley, CA, USA, in 1994 and 1995, respectively.

He was with the Rockwell International Laboratory on HBT modeling, where he developed the self-heating SPICE model for HBT. In 1996, he joined the ECE Department, The Hong Kong University of Science and Technology, Hong Kong, as a Faculty Member. In 2002, he joined the University of California, Berkley, as a Visiting Professor and the Co-Director of the BSIM project. His research at Berkeley covered a broad area in silicon devices ranging from process development to device design, characterization, and modeling. A major part of his work was on the development of recording breaking SOI technologies. He has also maintained a strong interest in device modeling and circuit simulation. He is one of the major contributors to the unified BSIM3 model for SPICE, which has recently been accepted by most U.S. companies and SEMATECH as an industrial standard model. He holds eight U.S. patents.

Prof. Chan was a recipient of the UC Regents Fellowship, the Golden Keys Scholarship for Academic Excellence, the SRC Inventor Recognition Award, the Rockwell Research Fellowship, the R&D 100 award (for the BSIM3v3 project), and other awards.