A Single-Ended Low Power 16-nm FinFET 6T SRAM Design With PDP Reduction Circuit

Chua-Chin Wang , *Senior Member, IEEE*, Ralph Gerard B. Sangalang , *Member, IEEE*, and I-Ting Tseng

Abstract—Memory arrays such as SRAM cells are responsible to the high-power consumption of modern digital systems. This investigation proposed an SRAM utilizing an ultra-low power cell, implemented using the 16-nm FinFET CMOS technology. Voltage supply selection of the static RAM cells is done by gating the wordline (WL) enable. In standby mode, the cell wordline is not activated, where the cell operates on a lower voltage level so that the stored bit status is still retained. On the other hand, the normal mode is activated when the wordline of the cell is enabled. Theoretical derivations, all-PVT-corner post-layout simulations, and measurement results were provided for verification of the functionality and performance. An SRAM of 1-kb capacity is designed based on the propose cell. The on-silicon measurement demonstrates 0.006832 fJ (energy/bit) at 500 MHz clock rate and 0.8 V supply.

Index Terms—Static RAM, built-in self test, voltage supply selector, PDP reduction circuit, read voltage boosting.

I. INTRODUCTION

MEMORY arrays that are made of Static Random Access Memory (SRAM) cells occupies a considerable chip area in modern microprocessors. It is also responsible to a significantly large power consumption of the chip. Prior reports stated that during charging and discharging of the bit-lines, on the read and write operations, dissipates up to 70% active power of the entire SRAM [1]. One way in decreasing power consumption is the usage of more advanced CMOS technologies to take advantage of its small size and lower operating voltages. However, planar CMOS technology nodes experiences problems such as, threshold voltage (V_{th}) reduction, degradation of the sub-threshold swing, drain-induced barrier reduction, fluctuations due to random channel dopant, and dielectric & band-to-band tunneling effects.

Many SRAM designs for the purpose of power and energy-saving have been proposed in the past decades. There are three major design approaches, namely, Current compensation [2],
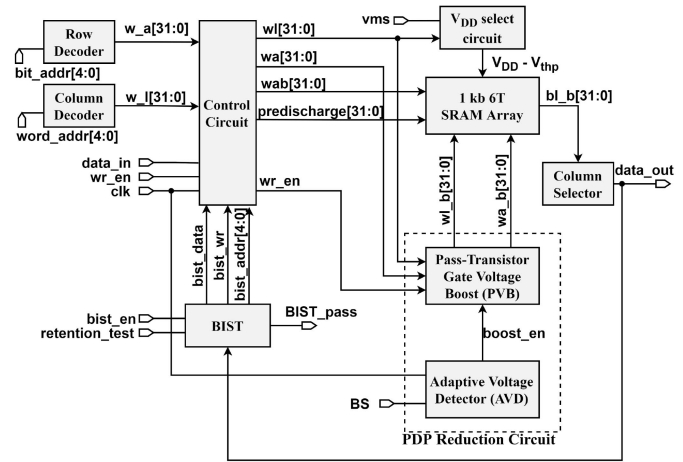
Fig. 1. Proposed SRAM System View.

Secondary supply [3], Current-mode sense amplification [4], and Split-cell supply [5]. However, they all suffered from different issues. Most of these challenges were somewhat relaxed upon the introduction of FinFET, since FinFET devices have several advantages over planar CMOS, including lower leakage power, higher speed performance, better mobility and scaling feature, a large drive current, and no random dopant fluctuation [6].

To have a lower power dissipation demand for SRAMs, a column gate-control mechanism is proposed in this investigation realized by the 16-nm FinFET technology. The word-line signal selects a lower supply voltage for cells that are idle thus decreasing the standby power dissipation. Validation of performance and functionality of the proposed design is justified on silicon.

II. LOW POWER SRAM DESIGN

The system view of the proposed SRAM is shown on Fig. 1, consisting of nine blocks, namely 1-kb SRAM Array, Row and Column decoder, Column multiplexer/selector, Controller, Built-In Self Test, V_{DD} Select Circuit, Pass-Transistor Gate Voltage Boost, and Adaptive Voltage Detector. The memory array is made up of 6T SRAM cells with power gating mechanism. The combination of adaptive voltage boost (AVB) circuit and pass-transistor voltage boost (PVB) circuit constitute the power delay product (PDP) reduction circuit.
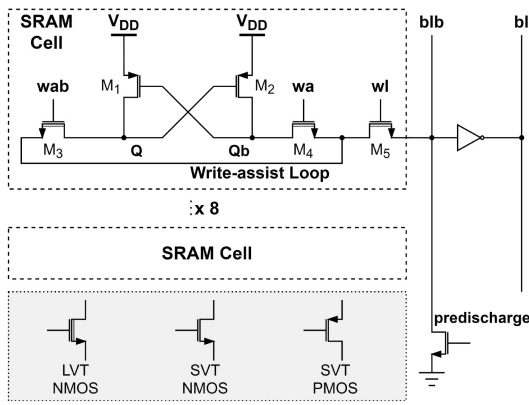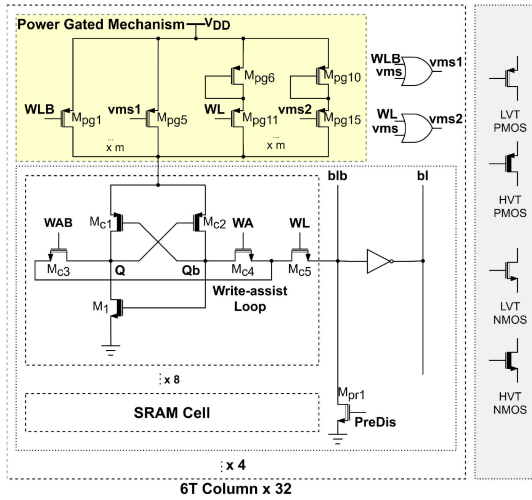
Fig. 2. 5T SRAM Cell [7].

Fig. 3. 6T SRAM Cell with Power Gating Mechanism.

A. Prior 5T Planar Cell Design

Fig. 2 as reported in [7] is a 5T loadless SRAM cell with a shared inverter. Two standard threshold voltage PMOS were used as a latch-like storage. To have a higher read and write current, low threshold voltage NMOS were used to access the storage. A low threshold voltage NMOS, M_5, is placed as an isolation to reduce the interference from blb (bitline_bar) to the cell. Though the isolation device exists, the low data state at node "Q" is at risk because of the write-assist loop leakage. With this, it might suffer from retention fault.

B. Proposed FinFET SRAM Cell

The mentioned retention problem, can be resolved by introducing a transistor at the foot of the latch, M_1 as shown in Fig. 3. When Qb is "high" (Q is "low"), the foot transistor will be on. Thus the low state at Q is ensured. Also, a high threshold voltage transistors for the latch is proposed. This is to make sure that the foot transistor will not be turned on due to the leakage that might be brought by the access transistors. Besides, high V_{th} PMOS devices were chosen to accommodate a new column structure to further lessen the power dissipation. The column structure of the cells is modified to improve power dissipation. The new structure is presented in Fig. 3

TABLE I
READ/WRITE OPERATION

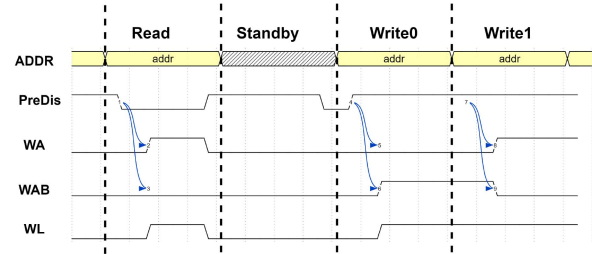|	Read (1/0)	Write-0	Write-1	Standby
PreDis	0	1	1	1
WA	1	1	0	0
WAB	0	0	1	0
WL	1	1	1	0
blb	0/1	0	0	0
bl	1/0	1	1	1

Fig. 4. Read and Write Cycle Timing.

based on [8]. The operation to reduce the standby power are the following.

a) M_{pg1} is on providing V_{DD} to the cells, when accessing a cell in a column.

b) A reduced voltage supply, $V_{DD} - V_{thp}$, is provided to the cells, if no cells are being accessed. This voltage drop will in turn save power.

C. Read and Write Cycles

The read/write cycles and associated control signals of the proposed design are presented in Table I. Read operation is described as follows.

● PreDis drives blb to the ground before any operation to prevent the low state disturbed by leakage and noise.

● The cells are selected by the corresponding decoders when the row & column addresses are ready.

● WA and WL are then set to "1" to turn on M_{c4} and M_{c5}. Regardless whether Read-1 or Read-0, WAB is "0" to turn M_{c3} off. Qb is now connected to blb through M_{c4} and M_{c5}.

The Write operation is as follow.

● *Write1:* WA is set to "1" to switch M_{c4} on, WAB is "0" to switch M_{c3} off. Pre-discharge pulls Qb down to pull Q up.

● *Write0:* WA is set to "0" and WAB is then set to "1" to switch M_{c4} and M_{c3} off and on, respectively. Q is then pulled down by the PreDis signal.

D. Built-in Self-Test (BIST)

Fig. 1 shows a BIST circuit included in Fig. 1 since it is required for high reliability in almost every memory system. The design uses a BIST circuit which was based on the March C-algorithm [9], featured with medium fault coverage and complexity. It was verified to detect at least stuck-at faults, data retention faults, transition faults, coupling faults, and address-decoder faults. The complexity of the March C-algorithm, which is one of the best BIST circuits, is 10×N, where N denotes the memory size, and it is 1024 in this investigation.
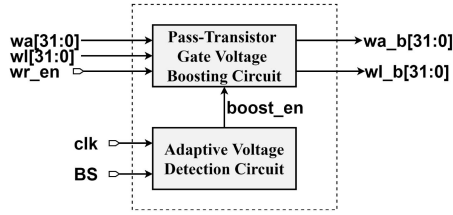
Fig. 5. PDP Reduction Circuit.

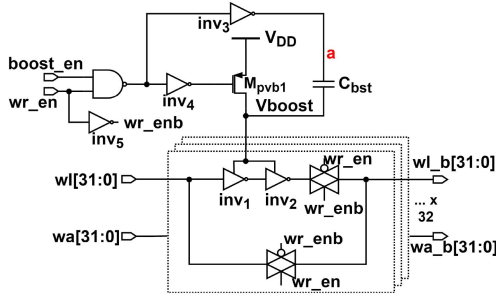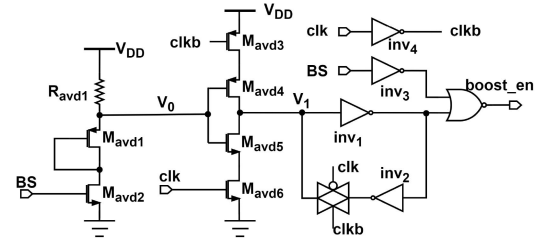Fig. 6. Pass-transistor Voltage Boost Circuit.
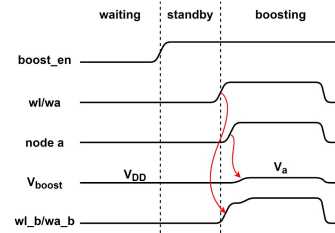

Fig. 7. Adaptive voltage detection circuit.

Fig. 8. Pass-transistor voltage boost timing diagram.

Fig. 9. Die photo of the proposed SRAM.
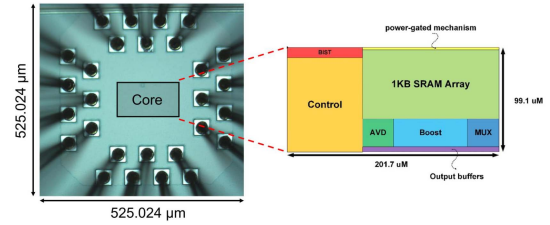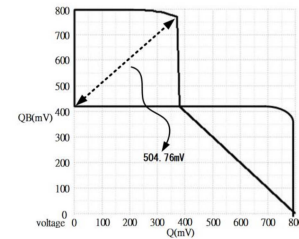

Fig. 10. Static noise margin (SNM) simulations.

The March C-algorithm is outlined as follows:

$$\{ \updownarrow (w0); \Uparrow (r0, w1); \Uparrow (r1, w0);$$
$$\Downarrow (r0, w1); \Downarrow (r1, w0); \updownarrow (r0) \} \qquad (1)$$

where $\Downarrow$ denotes down counts, $\Uparrow$ represents counting up, $\updownarrow$ is either up or down count, r represents read operation, and w means write access operation. The testing patterns are generated by a linear feedback shift register (LFSR) operating as a pseudo-random number generator with a characteristic equation as shown in Eqn. (2).

$$f(x) = x^5 + x^4 + x^3 + x + 1. \qquad (2)$$

E. Power Delay Product Reduction

Aside from using a power gating mechanism for each column of cells in the previous sections, power-delay product reduction (PDP) circuit is adopted for energy effective read/write operations. The PDP reduction circuit shown in Fig. 5 is composed of two sub-circuits: pass-transistor gate voltage boost circuit (PVB) and adaptive voltage detection (AVD) circuit [10]. The basic operation is that when boost select (BS) is "1", the AVD circuit generates the boost enable (boost_en) to the PVB circuit, thus giving the accessed cells a higher supply voltage V'_{DD} (a voltage higher than V_{DD}), for high-speed access operations.

1) Pass-Transistor Gate Voltage Boost (PVB) Circuit: The PDP reduction circuit is in waiting mode by default. This means that system voltage detection has not been completed by the adaptive voltage detector (V_0 vs. the switching threshold of the inverter in Fig. 7), the PDP reduction circuit will remain in a waiting mode. Once the PDP reduction circuit exits waiting mode, the circuit enters PDP reduction mode. The PVB circuit timing diagram is presented in Fig. 8.

2) Adaptive Voltage Detection: The adaptive voltage detector circuit used to produce the boost enable signal for the pass transistor voltage boosting circuit is seen in Fig. 7. The signal BS is fed to a common source amplifier to generate the signal VP0 which is then fed through the clocked inverter composed of M_{avd4} and M_{avd5} designed with a specific switching voltage to adapt to small changes with the BS signal. The output of the inverter is then latched to maintain a signal indicating a need to enable or disable the pass transistor gate boosting circuit. The boost_en signal will go high if the latched voltage V_1 and the output of inv_3 are both low.

III. SIMULATIONS AND MEASUREMENT

The proposed design is implemented using TSMC 16-nm FinFET technology. The die photo of the proposed SRAM is shown in Fig. 9. It has chip area of $525 \times 525 \ \mu m^2$ with a core area of $99.1 \times 201.7 \ \mu m^2$. The all-PVT-corner post-layout simulations were conducted. The worst-case static noise margin (SNM) is 504.76 mV, as shown in Fig. 10. Since the design

TABLE II
PERFORMANCE COMPARISON OF LOW POWER/ENERGY SRAM DESIGNS

	TVLSI	TVLSI	TCAS-I	TCAS-II	JCSC	SSCL	AICSP	**Ours**
Year	2016 [11]	2017 [12]	2017 [13]	2018 [14]	2019 [15]	2020 [16]	2021 [17]	**2021**
CMOS Tech. (nm)	40	65	65	65	28	40	65	**16**
Cell	5T	6T	9T	6T	6T	8T	6T	**6T**
Supply Volt. (V)	0.8	1.2	0.35	1.2	0.8	0.6	0.4	**0.8**
Verification^a	Meas.	Simu.	Meas.	Simu.	Meas.	Meas.	Meas.	**Meas.**
SNM (mV)	353	N/A	N/A	N/A	292	N/A	0.42	**504.76**
Read PDP (fJ)	N/A	17.5	N/A	N/A	444.5	N/A	N/A	**2.69**
Capacity (kb)	4+1	1	4	8	1+1	8096	16	**1**
Word Length	5	32	64	32	32	128	128	**32**
Frequency (MHz)	54	100	0.741	20	40	0.1	6	**500**
Energy/access (pJ)	0.941	2.2	0.229	0.592	0.0206	N/A	N/A	**0.00219**
Energy/bit (fJ)	188.22	68.75	3.58	18.5	0.6	67	0.536	**0.0068**
Core Area (mm^2)	0.024	0.013	0.011	0.019	0.025	0.082	0.01	**0.02**
FOM (fJ/V)^b	235.275	57.2917	10.2286	15.4167	0.75	111.6667	1.34	**0.0085**

^{a}Simulations (Simu.) or on-chip Measurements (Meas.) $\quad$ ^{b}FOM = (Energy/bit)/V_{DD}
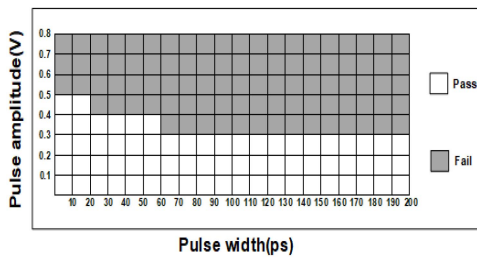
Fig. 11. Dynamic noise margin (DNM) simulations.

Fig. 12. Setup environment.

Fig. 13. Measured waveform of SRAM at 500 MHz (a) The assistant circuit is turned on. (b) The assistant circuit is turned off.

Fig. 14. Energy/bit technology road-map for SRAMs.
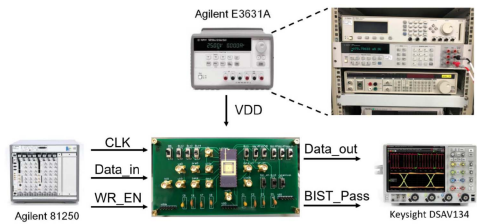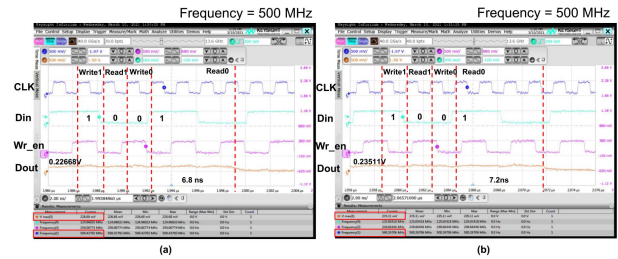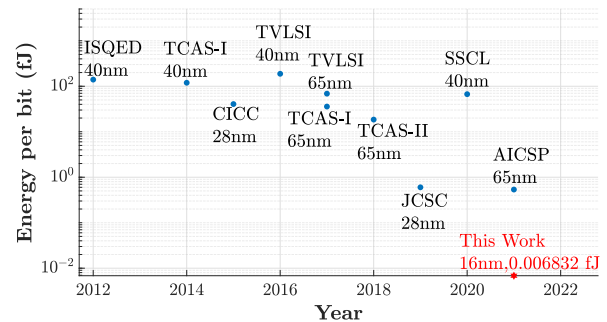
is implemented using a single-ended topology, the static noise margin is not the usual symmetrical butterfly shaped SNM as that of 2-bitline SRAM cells. The dynamic noise margin (DNM) shown in Fig. 11 indicates that the proposed SRAM cell can operate for as low as 300 mV.

Fig. 12 shows the measurement setup in Tainan Branch, TSRI (Taiwan Semiconductor Research Institute), where Agilent 81250 is a programmable function generator, and Keysight DSAV134 is the mixed-signal oscilloscope. A total of 8 chips were measured, where each one was measured 10 times (clock rate from 100 MHz to 1 GHz, 100 MHz step). Fig. 13 demonstrated the measured waveforms at clock rate = 500 MHz and V_{DD} = 0.8V. Fig. 13(a) and (b) are the results with and without PDP Reduction Circuit, respectively, where the delays are different. The circuit shows a 7.2-ns delay when the PDP reduction circuit is turned off, while it shows an improvement of 6.8-ns when the PDP reduction circuit is turned on.

As shown in Fig. 14, the proposed design achieved 0.0068 fJ energy per bit which is the lowest in the past 10 years as

compared to previous reports, as shown in Table II. Moreover, if the energy per bit is normalized to the supply voltage used, the proposed design shows the best results among all designs. As shown in Table II, the normalized value is 0.0085 fJ/bit/V.

IV. CONCLUSION

A very low power SRAM design implemented using 16-nm FinFET technology with power supply gating in response to the cell operation is presented in this investigation. Our design shows the best energy/bit by far thanks to the novel cell design, power gating circuit, and FinFET technology.

REFERENCES

[1] L. Villa, M. Zhang, and K. Asanovic, "Dynamic zero compression for cache energy reduction," in *Proc. 33rd Annu. IEEE/ACM Int. Symp. Microarchit. (MICRO)*, Monterey, CA, USA, Dec. 2000, pp. 214–220.

[2] K. Agawa, H. Hara, T. Takayanagi, and T. Kuroda, "A bitline leakage compensation scheme for low-voltage SRAMs," *IEEE J. Solid-State Circuits*, vol. 36, no. 5, pp. 726–734, May 2001.

[3] D. Kim, G. Chen, M. Fojtik, M. Seok, D. Blaauw, and D. Sylvester, "A 1.85fW/bit ultra low leakage 10T SRAM with speed compensation scheme," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Rio de Janeiro, Brazil, May 2011, pp. 69–72.

[4] A.-T. Do, Z.-H. Kong, K.-S. Yeo, and J. Y. S. Low, "Design and sensitivity analysis of a new current-mode sense amplifier for low-power SRAM," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 2, pp. 196–204, Feb. 2011.

[5] M. S. M. Siddiqui, Z. C. Lee, and T. T.-H. Kim, "A 16-kb 9T ultralow-voltage SRAM with column-based split cell-VSS, data-aware write-assist, and enhanced read sensing margin in 28-nm FDSOI," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 29, no. 10, pp. 1707–1719, Oct. 2021.

[6] C.-C. Wang and S.-W. Lu, "2.5 GHz data rate 2 × VDD digital output buffer design realized by 16-nm FinFET CMOS," in *Proc. 8th Int. Symp. Next Gener. Electron. (ISNE)*, Zhengzhou, China, Oct. 2019, pp. 1–3.

[7] D.-S. Wang, Y.-H. Su, and C.-C. Wang, "A readout circuit with cell output slew rate compensation for 5T single-ended 28 nm CMOS SRAM," *Microelectron. J.*, vol. 70, pp. 107–116, Dec. 2017.

[8] C.-C. Wang and I.-T. Tseng, "Ultra low power single-ended 6T SRAM using 40 nm CMOS technology," in *Proc. Int. Conf. IC Design Technol. (ICICDT)*, Suzhou, China, Jun. 2019, pp. 1–4.

[9] S. M. Al-Harbi and S. K. Gupta, "An efficient methodology for generating optimal and uniform march tests," in *Proc. 19th IEEE VLSI Test Symp. (VTS)*, Marina Del Rey, CA, USA, Apr. 2001, pp. 231–237.

[10] Y.-W. Chiu *et al.*, "40 nm bit-interleaving 12T subthreshold SRAM with data-aware write-assist," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 9, pp. 2578–2585, Sep. 2014.

[11] C.-C. Wang, D.-S. Wang, C.-H. Liao, and S.-Y. Chen, "A leakage compensation design for low supply voltage SRAM," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 5, pp. 1761–1769, May 2016.

[12] J. Lee, D. Shin, Y. Kim, and H.-J. Yoo, "A 17.5-fJ/bit energy-efficient analog SRAM for mixed-signal processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2714–2723, Oct. 2017.

[13] K. Shin, W. Choi, and J. Park, "Half-select free and bit-line sharing 9T SRAM for reliable supply voltage scaling," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 8, pp. 2036–2048, Aug. 2017.

[14] N. Surana and J. Mekie, "Energy efficient single-ended 6-T SRAM for multimedia applications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 6, pp. 1023–1027, Jun. 2019.

[15] C.-C. Wang, Z.-Y. Hou, D.-S. Wang, and C.-L. Hsieh, "A single-ended 28-nm CMOS 6T SRAM design with read-assist path and PDP reduction circuitry," *J. Circuits Syst. Comput.*, vol. 29, no. 6, Aug. 2020, Art. no. 2050095.

[16] J. Wang, H. An, Q. Zhang, H. S. Kim, D. Blaauw, and D. Sylvester, "A 40-nm ultra-low leakage voltage-stacked SRAM for intelligent IoT sensors," *IEEE Solid-State Circuits Lett.*, vol. 4, pp. 14–17, 2020.

[17] M. Nabavi and M. Sachdev, "A 350 mV, 2 MHz, 16-kb SRAM with programmable wordline boosting in the 65 nm CMOS technology," *Analog Integr. Circuits Signal Process.*, vol. 109, no. 1, pp. 213–224, Jul. 2021.