A 1.85fW/bit Ultra Low Leakage 10T SRAM with Speed Compensation Scheme

Daeyeon Kim, Gregory Chen, Matthew Fojtik, Mingoo Seok, David Blaauw, Dennis Sylvester
Department of Electrical Engineering and Computer
University of Michigan, Ann Arbor, MI USA
{daeyeonk, grgkchen, mfojtik, mgseok, blaauw, dmcs}@umich.edu

***Abstract*— A low leakage memory is an indispensable part of any sensor application that spends significant time in standby (sleep) mode. Although using high V_{th} (HVT) devices is the most straightforward way to reduce leakage, it also limits operation speed during active mode. In this paper, a low leakage 10T SRAM cell, which compensates for operation speed using a readily available secondary supply, is proposed in a 0.18µm CMOS process. It achieves the lowest-to-date leakage power consumption and achieves robust operation at low voltage without sacrificing operation speed. The 10T SRAM has a bit-cell area of 17.48µm^2 and is measured to consume 1.85fW per bit at 0.35V.**

I. Introduction

Sensors with long lifetime are becoming increasingly popular in areas such as medical, infrastructure, and environmental monitoring [1][2]. In sensor applications, reducing the standby power consumption is as important as reducing the active power consumption since the sensors spend significant time in standby mode. To minimize the standby power consumption, designing low leakage memory is indispensible [3][4][5][6]. Often, the leakage power consumption from memories dominates the total standby power consumption, since data stored in memory must be retained while most other blocks such as CPU, radios, and sensors can be fully power gated.

A low leakage 14T SRAM cell with stacked HVT devices [1] has been previously proposed; however, its area is 9.1× larger than the traditional 6T cell [7] and the HVT devices degrades write performance by ~10× compared to the read speed. To overcome these limitations and reduce leakage further, this work proposes a new ultra low leakage SRAM, referred to as the low leakage 10T SRAM, that exploits a boosted supply. We show how the boosted supply can increase operation speed and reduce leakage power simultaneously. Sensor applications typically operate using batteries, such as thin film batteries which tend to have high supply voltages. To obtain the subthreshold operating voltages, a common method for DC-DC conversion is to use a switched-capacitor networks (SCN) followed by a low-dropout regulator (LDO). In this case, boosted supply can be obtained with minimal overhead since it can be directly obtained from the input of the LDO or from a higher voltage output from the ladder SCN [2]. Also, several circuit techniques, including a floating bit-line scheme, word-line keeper, and read buffer, are introduced to reduce leakage further and guarantee robust read and write operation.

A prototype chip, which has 24kb of the low leakage 10T SRAM, shows that a bit-cell consumes 1.85fW of standby power at 0.35V with 0.5V of boosted supply. To our knowledge, this marks the lowest-to-date SRAM leakage power. The bit-cell area (Fig. 1) is 17.48µm^2, 3.97× larger than a traditional 6T cell [7] but 2.3× smaller than the previous low leakage 14T SRAM [1]. Since logic design rules are used in this design, area overhead can be mitigated with pushed SRAM design rules. This SRAM is successfully demonstrated as a part of an integrated sensor system with a CPU, power management unit, solar cells, and battery [2].

II. Low Leakage 10T SRAM Design

A. SRAM Bit-cell and Operation Modes

Fig. 2 shows the proposed 10T SRAM schematic to minimize leakage current without sacrificing operation speed. It consists of a 6T cross-coupled structure and a 4T read buffer. The read buffer can be power gated while the cross-coupled structure must remain on to retain data. Thus, standard V_{th} (SVT) devices are used in the

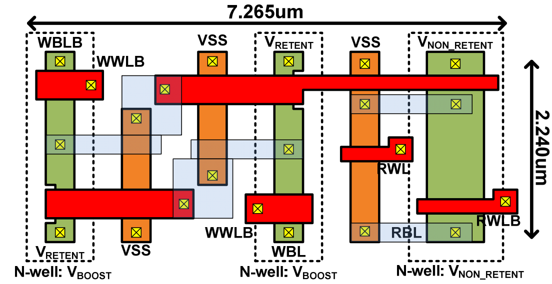

Figure 1. Logic design rules are used for the 10T SRAM layout.

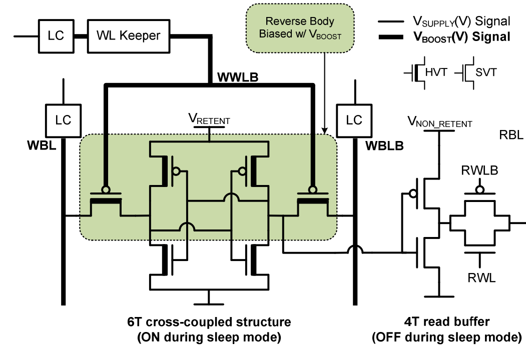

Figure 2. A low leakage 10T SRAM schematic is shown. Three signals (WBL, WBLB, WWLB) are boosted to V_{BOOST} using level converters to enhance write operation. Four PMOS devices in 6T cross-coupled structure are reverse body biased with V_{BOOST} for further leakage reduction.

read buffer for fast read operation and HVT devices are used in the cross-coupled structure for minimizing leakage. The layout is shown in Fig. 1 and logic design rules are used.

This SRAM operates with three different power supplies: V_{RETENT}, V_{NON_RETENT}, and V_{BOOST}. V_{RETENT} and V_{NON_RETENT} have the same voltage level but are connected to different power gating switches. V_{BOOST} has a higher voltage level than the other two supplies and is used for boosting bit-lines and reverse body biases the four HVT PMOS devices. Boosting bit-lines enhances write operation speed, and reverse body biasing allows further leakage reduction. The cell is still functional if V_{BOOST} is the same as V_{RETENT} and V_{NON_RETENT}.

There are three operation modes (See Table I). During active mode, V_{RETENT} and V_{NON_RETENT} are at V_{SUPPLY} while V_{BOOST} is higher than the two others. When the power gating switch connected to V_{NON_RETENT} is turned off, the system moves to standby mode. To retain data, V_{RETENT} still remains at V_{SUPPLY}. V_{BOOST} must also be kept on to turn off the access transistors and bias the n-well. If no data retention is required, all supplies can be turned off.

B. Bit-line Boosting for Fast Write Operation

The read operation already has an acceptable access time of below 20 SVT FO4 (Fan-out 4) delays including cascaded read buffer delays because SVT devices are used in the 4T read buffer. However, without further modification, the write operation limits the performance of this SRAM cell at more than 1000 SVT FO4 delays because of slow HVT devices in the 6T cross-coupled structure.

Write speed can be improved by increasing the ON current of the access transistor. First, PMOS access transistors are used instead of the traditional NMOS access transistors since, at low voltage in this technology, HVT PMOS devices have larger ON current than HVT NMOS devices. Second, word line boosting is adopted. With NMOS access transistors, writing "0" is dominant and a boosted word-line can increase ON current by raising V_{gs} of the NMOS (Fig. 3(b)). On the other hand, with PMOS access transistors, writing "1" is dominant and bit-line boosting is applied. With bit-line boosting, both V_{gs} and V_{ds} of the PMOS are boosted and therefore it results in better performance improvement. Since bit-line boosting is more effective than word-line boosting and since a negative power supply is not readily available, PMOS devices were selected.

Simulated behavior in Fig. 4(a) depicts HVT SRAM write speed improvement as boosted supply increases. The write speed in this plot does not include peripherals to directly compare the effect of bit-line boosting. The effect of reverse body biasing will be discussed in the following section. Without boosting, HVT SRAM needs more than 1000 SVT FO4 delays and, therefore, a processor must run many cycles for a single write operation. As boosted supply increases, the write speed is dramatically improved and the write operation can be executed in a single or a few cycles.

With bit-line boosting, level converters for word-lines are

TABLE I. SRAM OPERATION MODES

Mode / Voltage	Active	Standby	Shutdown
V_{RETENT}	V_{SUPPLY}	V_{SUPPLY}	0
V_{NON_RETENT}	V_{SUPPLY}	0	0
V_{BOOST}	$\geq V_{SUPPLY}$	$\geq V_{SUPPLY}$	0

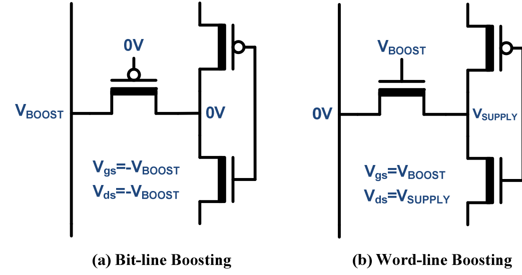

Figure 3. (a) Bit-line boosting with PMOS access transistor. (b) Word-line boosting with NMOS access transistor.

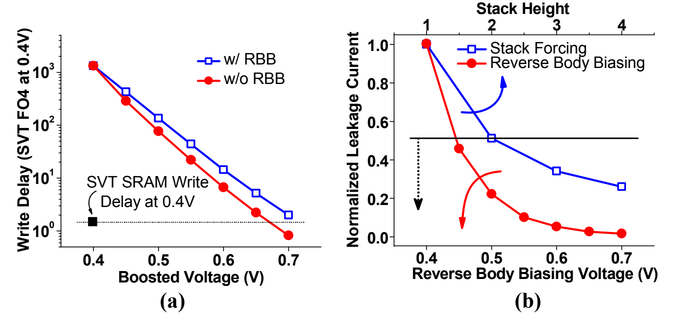

Figure 4. (a) Bit-line boosting significantly improves write speed (simulated results). (b) Leakage current of a PMOS device is shown. Stack height means the number of devices in a stack. Reverse body biasing is more effective than stack forcing (simulated results).

required. This is because unselected access transistors in the same bit-line must be fully turned off to prevent data corruption, requiring word-lines kept at V_{BOOST} during a write operation.

C. Body Biasing with Boosted Supply

Reverse body biasing of four HVT PMOS devices in the 6T structure is adopted for leakage minimization without increasing bit-cell area. Fig. 4(b) compares leakage current of an HVT PMOS device with stack forcing and reverse body biasing at 0.4V in simulation. It shows that stack forcing is not as effective as reverse body biasing at low voltage. With more than 50mV of reverse body biasing, leakage reduction is better than stacking two devices. In addition, stack forcing needs more devices and, therefore, increases bit-cell area. Because of the optimized layout of the 6T structure, adding stacked devices causes a more than 2× area increase.

Reverse body biasing decreases both ON current and OFF current, so it can degrade write operation significantly. However, the access transistor in the dominant writing "1" path does not experience reverse body biasing during write since the bit-line boosting scheme increases voltage level of source while the body biasing increases by the same amount. V_{bs} is still 0V during write operation, so reverse body biasing for all four HVT PMOS devices in the 6T structure does not weaken write operation. In Fig. 4(a), HVT SRAM can achieve sufficient speed improvement even with reverse body biasing.

D. Leakage Reduction During Standby Mode

There are four different leakage paths in an SRAM cell during standby mode (Fig. 5(a)). The 4T read path is power gated so it is not considered as a leakage path. V_{BIT} and $_{VBIT_B}$ affect I_{AXL} and I_{AXR}, but do not impact I_{PU} and I_{PD}. If V_{BIT} and V_{BIT_B} keep either V_{SUPPLY} or 0V, leakage current exists in only one path between I_{AXR} and I_{AXL} and the amount of leakage through each path is the same. This implies that the total leakage current does

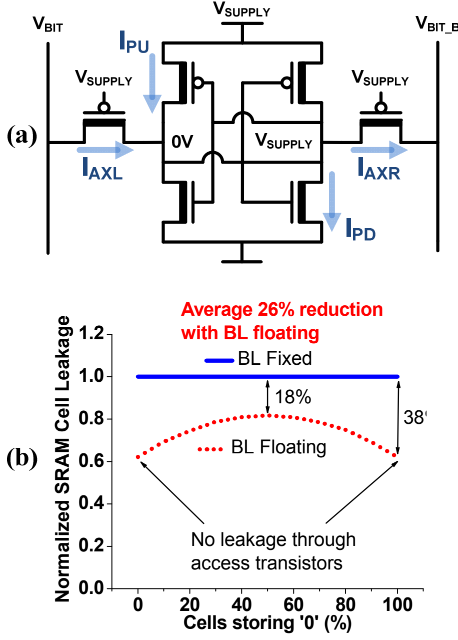

Figure 5. (a) There are four leakage paths during standby mode. (b) Bit-line floating shows at least 18% leakage reduction (simulated results).

not change as long as bit-lines are driven to V_{SUPPLY} or 0V. We propose using bit-lines that are floating. In this case, the voltage levels of bit-lines are determined by data stored in the cells connected to the same bit-line. With all the same data in a bit-line, there is no leakage through access transistors. Otherwise, V_{BIT} and V_{BIT_B} are in an intermediate voltage between V_{SUPPLY} and 0V and therefore an access transistor whose internal node is "0" is super cut-off. In simulation, this mechanism allows at least an addition 18% leakage reduction, and could decrease leakage further depending on the data stored in the cells (Fig. 5(b)).

Simulated results show that PMOS reverse body biasing is also effective to minimize leakage during standby mode (Fig. 6(a)). I_{AXL} and I_{AXR} can be minimized with BL floating and RBB, while I_{PU} can be minimized with RBB only. If two techniques are applied, the only remaining leakage path is I_{PD}. NMOS reverse body biasing to reduce I_{PD} is not practical since a triple well process is required, increasing bit-cell area tremendously. Additionally, it is relatively difficult to obtain negative power supply compared to boosted power supply since the boosted supply can be easily obtained from the higher voltage output from the ladder SCN in a DC-DC converter.

A special purpose word-line keeper (Fig. 6(b)) is designed to obtain two goals: no speed degradation and no SVT leakage path. The word-line keeper operates just like a normal word-line driver during active mode, but its output must be kept high in standby mode to fully turn off PMOS access transistors and prevent data corruption. The voltage level of SLEEP_B is higher than 3V (output voltage of small form-factor battery such as a Li battery) since this control signal is generated to control power gating switches. The use of the battery voltage level signal is justified since a small form-factor battery is used in most sensor applications.

E. Read Buffer Design

An improved 4T read buffer is designed for robust and fast read at low voltage. A static logic circuit (4T read buffer) can prevent erroneous bit-line discharge, which may occur in a dynamic logic circuit (2T read buffer [8]), due to its relatively small ON-OFF current ratio at low voltage. A clocked-gate type 4T read buffer was used in [1] (Fig. 7(a) Type 1) while a tri-state buffer type 4T read buffer (Fig. 7(a) Type 2) is used in this design. Type 2 is faster than Type 1 since both NMOS and PMOS in Type 2 can drive RBL when only one device can drive RBL in Type 1. With this new 4T read buffer, RBLs are cascaded instead of directly connecting all bit-cells to one global RBL. In RBL cascading, eight bit-cells are connected to a local RBL and then local RBLs are cascaded to a global RBL. In the worst case, data in the all unselected cells are different from data in the targeted cell and therefore RBL leakage disturbs read operation. Since RBL leakage can be minimized with cascading, it can improve read speed. With new 4T read buffer with cascading, read speed can be improved by 72% in simulation (Fig. 7(b)).

III. MEASURED RESULTS

A 32kb low leakage 10T SRAM was fabricated in a 0.18µm CMOS process with nominal voltages of 1.8V and 3.3V for SVT and HVT respectively.

The SRAM array has 768 words and each word has 32-bit data. In TABLE II, the first 32 words are tested to measure how many words fail as supply and operating frequency are swept. This table shows that there are more words that fail at low voltage and high frequency.

Fig. 8 depicts speed improvements as boosted supply increases. At 0.35V, the whole SRAM array (768 words) with peripherals operates at 3.5 kHz without read and write fail. If the boosted supply is applied, the write speed is enhanced and therefore the system can operate substantially faster. With 0.5V of boosted supply, the operating frequency can reach 52.5kHz, which is ~185 SVT FO4 delays at 0.35V, and this is 15× speed improvement. However, the speed improvement is saturated since a read buffer, peripherals, and CPU still run at 0.35V and the

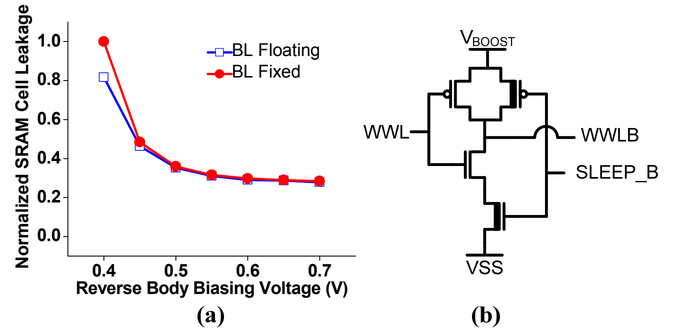

Figure 6 (a) PMOS reverse body biasing (simulated result). (b) Word-line keeper design.

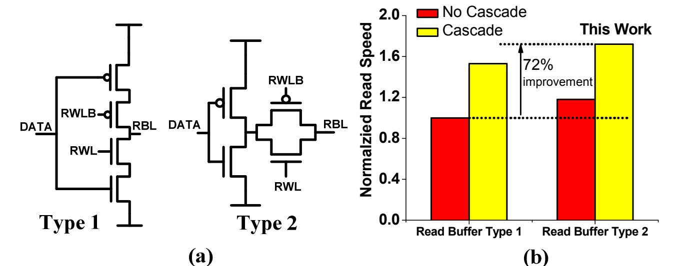

Figure 7. (a) Two types of 4T read buffer. (b) Read buffer type 2 with cascade increases read speed 72% (simulated result).

TABLE II. NUMBER OF READ/WRITE FAILURE WORDS

Supply(V)	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
0.3V	0	23	32	32	32
0.325V	0	0	31	32	32
0.35V	0	0	1	32	32
0.375V	0	0	0	3	32
0.4V	0	0	0	0	12

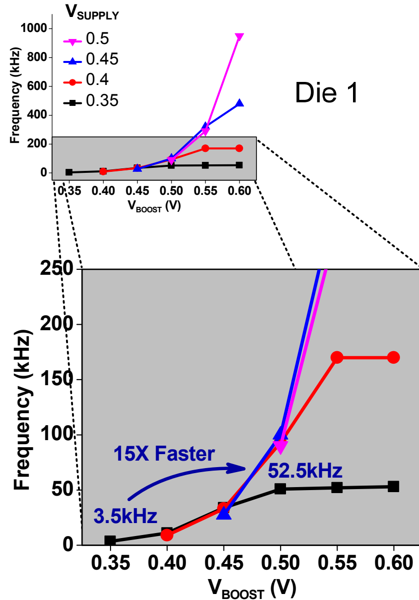

Figure 8. Operation speed is improved 15X with boosted supply (measured result).

boosted supply does not change their operation speed.

Fig. 9(a) shows the measured leakage power per bit-cell as supply and boosted supply are swept in two different dies. Without boosted supply, the total leakage power monotonically increases as supply increases. If boosted supply is applied, the leakage power through V_{SUPPLY} significantly decreases as it reverse body biases PMOS devices. However, the leakage power through V_{BOOST} increases since it includes leakage through word-line keeper as well as body leakage. Because of these two different power trends, the optimal power minimum (P_{min}) point can be found. Fig. 9(b) shows the power breakdown. Two dies in Fig. 9(a) show almost equivalent leakage reduction characteristics. At 0.35V (Fig. 9(b)), the initial leakage power is 3.6fW without boosting, but it can be minimized down to 1.85fW with 0.5V of boosted supply. This is 49% leakage power reduction. In Fig. 10(a), leakage powers through V_{RETENT} are normalized in the different temperatures at 0.35V. In the different temperature, the boosted supply still allows large leakage reduction. Fig. 10(b) shows the chip micrograph and dimension.

IV. CONCLUSION

A low leakage 10T SRAM which consumes femtowatt-scale leakage power is proposed for long lifetime sensor applications. A boosted supply is exploited to compensate slow write operation caused by HVT devices and minimize leakage further. A prototype chip fabricated in a 0.18µm CMOS process shows that 1.85fW of leakage power at 0.35V and operates at 52.5kHz with boosted supply. The boosted supply allows 49% of leakage power reduction and 15× speed improvement at 0.35V.

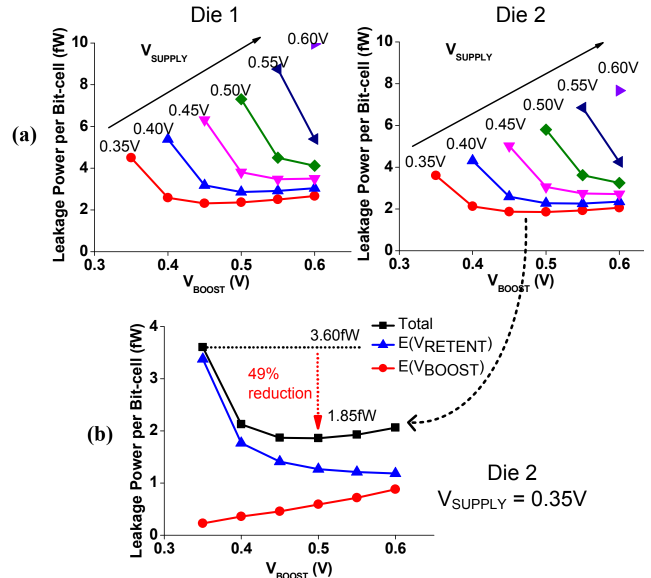

Figure 9. (a) Total leakage power is shown as V_{SUPPLY} and V_{BOOST} are swept in two different dies (measured result). (b) Leakage power can be minimized with boosted supply (measured result).

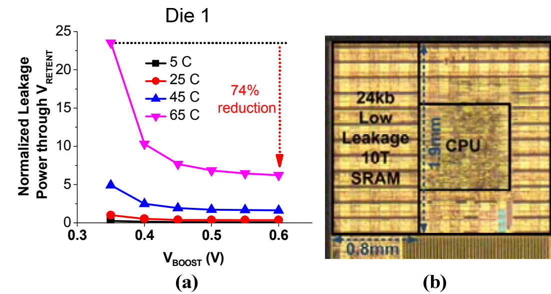

Figure 10. (a) Temperature variation of normalized leakage through V_{RETENT} (measured result). (b) Chip micrograph and dimension in 0.18µm CMOS process

REFERENCES

[1] S. Hanson et al., "A Low-Voltage Processor for Sensing Applications with Picowatt Standby Mode," IEEE Journal of Solid State Circuits, Vol. 44, pp. 1145-1155, Apr. 2009.

[2] G. Chen et al., "Millimeter-Scale Nearly Perpetual Sensor System with Stacked Battery and Solar Cells," IEEE International Solid-State Circuits Conference, pp. 288-289, 2010.

[3] T. Kim, J. Liu, C. Kim, "A Voltage Scalable 0.26V, 64kb 8T SRAM with Vmin Lowering Techniques and Deep Sleep Mode," IEEE Journal of Solid State Circuits, Vol. 44, pp. 1785-1795, June 2009.

[4] T. Kim, J. Liu, J. Keane, C. Kim, "A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing," IEEE Journal of Solid State Circuits, Vol. 43, pp.518-529, Feb. 2008.

[5] N. Verma, A. P. Chandrakasan, "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," IEEE Journal of Solid State Circuits, Vol. 43, pp.141-149, Jan. 2008.

[6] H. Mair et al., "A 65-nm Mobile Multimedia Applications Processor with an Adaptive Power Management Scheme to Compensate for Variations," IEEE Symposium on VLSI Circuits, pp. 224-225, 2007

[7] C. H. Diaz et al., "A 0.18 µm CMOS Logic Technology with Dual Gate Oxide and Low-k Interconnect for High-Performance and Low-Power Applications," Symposium on VLSI Technology, pp. 11–12, June 1999.

[8] L. Chang et al., "Stable SRAM Cell Design for the 32nm Node and Beyond," IEEE Symposium on VLSI Circuits, pp. 128-129, 2005