Received 15 February 2021; revised 7 June 2021 and 1 August 2021; accepted 5 August 2021. Date of publication 17 August 2021; date of current version 6 September 2021.

Digital Object Identifier 10.1109/OJCAS.2021.3104945

Ultralow-Voltage Retention SRAM With a Power Gating Cell Architecture Using Header and Footer Power-Switches

HAYATO YOSHIDA, YUSAKU SHIOTSU (Graduate Student Member, IEEE), DAIKI KITAGATA (Member, IEEE), SHUU'ICHIROU YAMAMOTO (Member, IEEE), AND SATOSHI SUGAHARA (Member, IEEE)

Laboratory for Future Interdisciplinary Research of Science and Technology, Tokyo Institute of Technology, Yokohama 226-8502, Japan

This article was recommended by Associate Editor J. Park.

CORRESPONDING AUTHOR: Y. SHIOTSU (e-mail: y.shiotsu@isl.titech.ac.jp)

This work was supported by VLSI Design and Education Center (VDEC), The University of Tokyo, in collaboration with NIHON SYNOPSYS G.K and with Renesas Electronics Corp.

ABSTRACT An ultralow-voltage retention SRAM (ULVR-SRAM) cell using header and footer power-switches (HFPSs) is investigated for power-gating (PG) applications. The cell can change its operational mode depending on the cell voltage ($V_{\rm cell}$) controlled by the HFPSs: When the ordinary supply voltage is applied, the cell can act as a high-performance SRAM cell. When $V_{\rm cell}$ is reduced to an ultralow voltage, the cell can transition to the ULVR mode and dramatically reduce the leakage power without losing its data, i.e., the substantive PG can be achieved using the ULVR. The ability of leakage power reduction is enhanced by introducing the body biases that are automatically induced only during the ULVR mode. The design methodology is developed based on quasi-static noise margins, where the transistor sizes and the bias condition for $V_{\rm cell}$ are determined so as to minimize the leakage power with keeping a sufficiently high noise margin in the ULVR mode. An optimally designed ULVR-SRAM cell shows excellent PG ability: The leakage power can be reduced by ~98% using the ULVR and a minimally short break-even time of 1.5μs can be achieved for the 8KB macro. The ULVR-SRAM can provide a new class of energy efficient PG architecture.

INDEX TERMS SRAM, cell design, ultralow voltage retention, power gating, standby power, break-even time, noise margin.

I. INTRODUCTION

STATIC random access memory (SRAM) embedded in CMOS logic systems such as microprocessors (MPs) and system-on-chip devices (SoCs) plays an essential role as a basis for computing architectures [1]. Furthermore, it strongly affects the speed and power performances of these logic systems as a key factor [2]–[5]. In emerging neural network accelerators, SRAM is also the indispensable component that governs their information processing ability [6]–[8]. Standby power (static leakage power) in SRAM due to thermally activated leakage currents has been increasing with scaling-down transistor sizes [9], [10], which has a great impact on logic system designs.

Power gating (PG) is an effective technique to reduce standby power in CMOS logic systems [11]–[13], and current standard MPs and SoCs implement this leakage power reduction technique [14], [15]. In PG architectures, a logic system is partitioned into functionally/constitutionally meaningful circuits, i.e., the so-called power domains. The power management of these power domains is carried out with the aid of power switches (PSs) installed to them, which can shut off the power supply to standby-state domains, resulting in effective reduction of the standby power. However, there are great challenges facing PG architectures, originating in various memory circuits (such as caches) configured with SRAMs. These SRAM-based memory circuits cannot shut down without losing their data owing to the volatile nature. Therefore, sleep (moderately-low-voltage retention) and data flush techniques are employed for leakage reduction of SRAM-based memory circuits [3], [4], [14], [15]. However, these are not

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

necessarily sufficient to reduce the standby power owing to lower power-reduction efficiency (for the former) or larger temporal granularity for PG executions (for the latter).

Emerging nonvolatile memories such as magneto-resistive random access memory (MRAM) [16], [17] and nonvolatile SRAMs (NV-SRAMs) using nonvolatile memory elements [18], [19] have attracted considerable attention for PG applications [20], [21]. However, the overheads of energy and latency due to the write operation to the nonvolatile memory elements restrict the energy reduction efficiency (or granularity of PG). Furthermore, the embedded process of nonvolatile memory elements costs a great deal.

Ultralow-voltage retention (ULVR) using fully CMOS-based SRAM is an alternative technology to achieve efficient PG for memory circuits in logic systems. Recently, an ULVR-SRAM cell (ULVR cell) is proposed toward this end [22], which is comprised of Schmitt-trigger-based dual-mode inverters that can change its operational mode depending on the applied voltage. The cell can act as a high performance SRAM cell for the ordinary-supply-voltage operations. When the supply voltage is reduced to an ultralow voltage (ULV), the cell can transition to the ULVR mode and dramatically reduce the standby power without losing its data. However, there is a trade-off relation between the leakage current reduction and noise margin securement of the cell during the ULVR mode, i.e., reduction of ULV causes both decreases in the leakage current and noise margin of the cell. A PG architecture using header and footer power-switches (HFPSs) has a possibility to effectively diminish the leakage current, since sufficient reverse body biases can be automatically applied to both the nMOS and pMOS devices in the cell during the ULVR mode. This would also weaken the trade-off relation, resulting in lower leakage power with a wider retention noise margin during the ULVR mode.

In this paper, we investigate a new ULVR cell architecture using HFPSs, in which the effective body bias is automatically induced in the cell only during the ULVR mode. The design methodology of the cell is developed with careful consideration of the device variability for maximizing the retention stability and minimizing the leakage current during the ULVR mode. The optimally designed ULVR cell exhibits excellent standby-power reduction ability and a minimally short break-even time (BET). The proposed ULVR-SRAM can provide a new class of power-gating architecture using the ULVR, whose energy reduction ability is comparable to PG architectures employing nonvolatile retention.

II. RELATED WORK

Firstly, PG techniques for SRAM and low-voltage SRAM technologies are briefly reviewed. PG architectures for SRAM utilize power switches (PSs) for controlling the supply voltage as those for logic circuits. There are two types of PS usages in PG architectures for SRAM: complete power-shutdown of the cell array and power-regulation to the cell array. For instance, these concrete architectures of the former and latter are so-called data flush and sleep (moderately-low-voltage retention) techniques, respectively. Recently, in many cases, both the techniques are implemented in MPs and SoCs [3], [4], [14], [15]. In the complete power-shutdown technique, data stored in the cells are lost by PG executions. Thus, chance for executing PG is limited, i.e., shortening the temporal granularity (enhancing the PG efficiency) is not easy. In this case, particular/special architectures for according data loss are required for efficient PG executions.

The leakage-power reduction ability of the power-regulation technique is not so high as that of the complete power-shutdown technique. Nevertheless, this technique can reduce the leakage power without losing data. In addition, the system can quickly return to the normal operation mode, and thus this technique can achieve frequent PG executions. A variety of architectures on PSs has been proposed for the power-regulation technique, including proper design of PSs [23], diode clamp [24], dual power rails with normal and low voltages [25], and autonomous active clamp using an operational amplifier [26].

In these architectures, the bit cell is comprised of a bistable circuit with an ordinary inverter loop, and thus the retention mechanism is equivalent for these architectures. The minimum retention voltage for such bistable circuits depends on the size of constituent transistors [26], PS design/configuration (header, footer, or both PSs), and retention method (voltage dividing, diode clamping, and low-voltage biasing). The low-voltage biasing is more useful than the diode clamping, since it can provide more design flexibility. In this technique, an architecture using both the header and footer PSs (HFPSs) is beneficial in comparison with only a header PS (HPS) or footer PS (FPS) architecture, as shown in this paper. The HFPS architecture can effectively reduce the leakage currents in the cell, which can also enhance the noise margin during the low-voltage retention mode owing to eliminating the instability caused by the leakage currents. The HFPS architecture can also be applied to automatic body-bias control [27], [28]. Although this technique was originally proposed for dynamic threshold voltage control, it can be used for highly effective leakage reduction in PG, as shown in this paper.

Low-voltage SRAM cells can be classified into several types. A typical cell can be configured with a 6T-based cell adding an independent READ port (IRP) [29]–[33] and/or a contention-free WRITE (CFW) circuit [34], [35]. Another strategy is to employ a bistable circuit consisting of Schmitt-trigger (ST) inverters [36], [37]. These technologies have been developed for low-voltage SRAM operations rather than for PG applications. For low-voltage SRAM operations, ensuring the noise margin of READ operation is severe. The IRP cells can discharge the bit line through the additional READ port isolated from the bistable circuit. Thus, the cell can achieve a sufficient margin even for the low-voltage READ operation. Although this type of cell basically has an 8T configuration [29], [30], several variations including 10T

cells have been proposed [31]–[33] (One of the purposes is to suppress a leakage current through the additional READ port). In this technique, the retention ability itself cannot be improved, since the bistable circuit part is the same as that of conventional 6T cells. Therefore, the minimum retention voltage is equivalent to the 6T cells.

The CFW cells also have several variations, such as 8T and 10T cells [34], [35]. In addition to the IRP, these cells are designed to achieve secure WRITE operation at low voltages. This type of cell generally has a single-ended WRITE port with a CMOS transfer gate, and the inverter connected to the WRITE port can be electrically isolated during the WRITE operation, resulting in the CFW operation. The basic idea is based on an architecture shown in [38]. The CFW cells also have the same retention mechanism as the conventional 6T cells. Although the IRP and CFW cells have been developed for low-voltage SRAM operations, they can also be applied to PG. However, the minimum retention voltage for PG cannot be reduced by the usage of these cells.

ST-inverter-based SRAM cells (ST cells) can also be applied to low-voltage operations [36], [37]. The ST inverters show rectangular transfer characteristics with hysteresis. As a result, the ST cells can improve the retention stability at low voltages. This feature ensures to enlarge the noise margin of the READ operation at low voltages. In addition, the minimum retention voltage can be reduced to further low voltages, i.e., ultra-low voltage retention (ULVR) is possible for the ST cells. Although conventional ST inverters have feedback transistors (FBTs) on both the pull-up and pull-down sides, a simple configuration using a one-side FBT is also possible, i.e., the number of transistors can be reduced in the ST inverter. Thus, a 10T ST cell can be realized [36]. These ST cells have a potential to achieve the ULVR that is applicable to efficient PG architectures. However, the leakage power significantly increases during the normal SRAM operation mode under the ordinary supply voltage. The ST cells are suitable for the exclusive use of low-voltage operations.

From the above discussions, size-adjusted/-enlarged 6T and ST cells are considered to be promising for the ULVR operation, although these cells need to reduce the leakage power during the ordinary-supply-voltage operation mode. The ST mode of the ULVR cell described in the previous section employs the same ULVR mechanism as the ST cell. In addition, the leakage current under the ordinary supply voltage can be suppressed using the normal inverter (NI) mode. Meanwhile, the ULVR based on the HFPS architecture with the automatic body-bias control is further promising for efficient leakage reduction. The strategy of this paper is to clarify the feasibility of PG using the ULVR for this new type of ST cell, i.e., the ULVR cell employing the HFPS architecture with the automatic body-bias control, which is systematically compared with the other candidates of size-adjusted 6T and conventional ST cells.

III. CELL ARCHITECTURE

Fig. 1 (a) shows the ULVR cell with HFPSs. In the figure, VV_{DD} and VV_{SS} represent the voltages of the virtual power supply and virtual ground rails, respectively (these notations are also used as the names of these rails). VV_{DD} is generated through the power switches PS_1 and PS_2 from two supply voltages V_{DDH} (= 1.2 V) and V_{DDL} (= 0.50 V). VV_{SS} is also generated using the power switches PS_3 and the PS_4 with V_{SSH} (= 0.30 V) and V_{SSL} (= 0.0 V). The net voltage V_{cell} applied to the cell is given by $V_{cell} = VV_{DD} - VV_{SS}$. In the cases of the normal SRAM operation and ULVR modes, V_{cell} is set to $V_{cellH} = V_{DDH} - V_{SSL}$ (= 1.2 V) and $V_{cellL} = V_{DDL} - V_{SSH}$ (= 0.20 V), respectively, using these PSs. Note that Fig. 1 also shows the body bias connections for these PSs. This biasing technique is effective to avoid the back-flow leakage through PS_2 (PS_4) from the V_{DDH} to V_{DDL} (V_{SSH} to V_{SSL}) power rails.

The ULVR cell is configured with ST-based dual-mode (DM) inverters, which can switch its operating mode by controlling the bias voltage (V_{FB}) applied to the cell. The DM inverter has the pMOS FBT whose bias can be changed through V_{FB} (= V_{SSL} or V_{DDL}), and its gate is connected to the output of the other inverter in the cell. The control (CTRL) driver is used for controlling V_{FB}. During the normal SRAM operation and ULVR modes, V_{FB} is biased to V_{SSL} and V_{DDL}, respectively. Thus, the DM inverters act as a normal inverter (NI) and a Schmitt trigger (ST) inverter during the normal SRAM operation mode ($V_{cell} = V_{cellH}$) and the ULVR mode ($V_{cell} = V_{cellL}$), respectively. As a result, the ULVR cell operates as a conventional SRAM cell at $V_{cell} = V_{cellH}$, and the cell can retain its data with a high noise margin based on the ST mode of the DM inverters even at $V_{cell} = V_{cellL}$. Note that in the cell array, the multiple cells can share with a set of the PSs and a CTRL driver that are arranged and integrated as peripheral circuits. Therefore, the cell can be configured as a 10T cell. The area overheads for the PSs and CTRL drivers are evaluated individually as peripherals, as discussed later.

The bodies of pMOS and nMOS devices are connected to the V_{DDH} and V_{SSL} rails, respectively. This connection configuration results in the automatic body-bias control for reducing the leakage currents only during the ULVR mode. The load pMOS and bottom driver nMOS devices are subject to the body biases of $V_{DDH} - V_{DDL}$ and $V_{SSL} - V_{SSH}$, respectively. This body-bias effect does not appear during the normal SRAM operation mode, meaning that the body bias effect has no performance degradation for the normal SRAM operations. Note that although a triple well is required to control the body biases of the ULVR cell, the cell array can be implemented in a single deep well.

It is worthy to note that a ULVR-SRAM cell can be also configured using nMOS FBTs [22]. Although the circuit configuration of the constituent inverters is completely identical with that of a general ST inverter [36], the bias of the nMOS FBTs can be independently controlled for the ULVR-SRAM

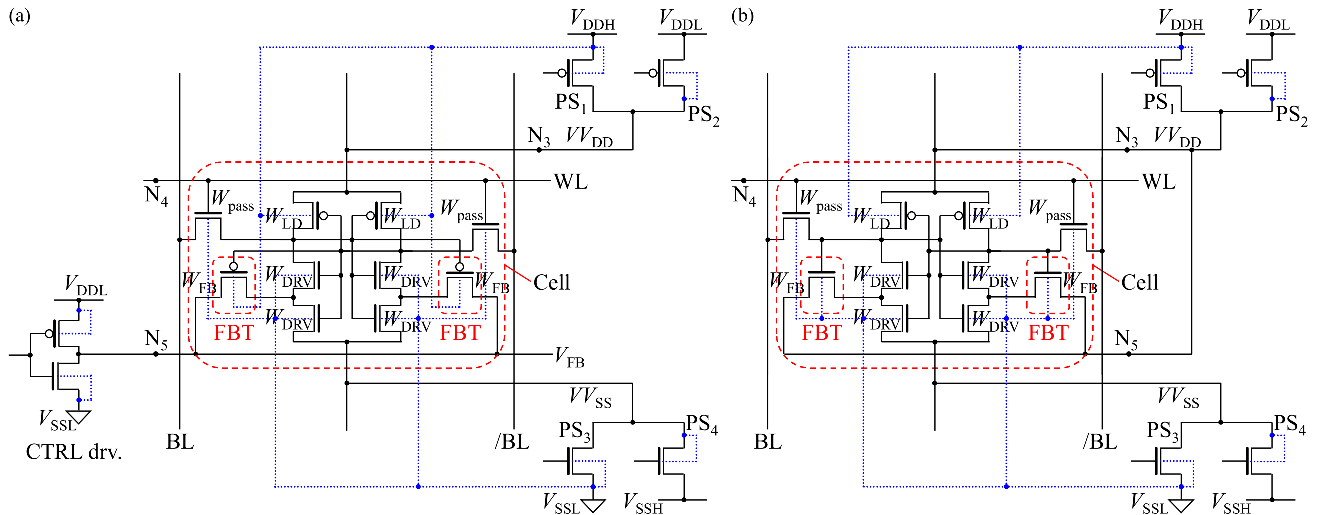

FIGURE 1. Circuit configurations of (a) ULVR- and (b) ST-SRAM cells with HFPSs. The dotted lines represent the connections for the automatic body bias control of the cells.

cell. This type of ULVR-SRAM cell suffers from the cell layout optimization for minimizing its area, which originates in the different numbers of constituent pMOS and nMOS transistors in the cell [22], [36]. On the other hand, the ULVR cell using the pMOS FBTs can easily minimize the cell area, as shown later.

Figs. 2 (a) and (b) show voltage transfer characteristics (VTCs) of the conventional, ST and DM inverters at $V_{\text{cell}} = 1.2$ V and 0.20 V, respectively, in which the after-mentioned designs are used for these inverters. The ST inverter exhibits the abrupt VTC with hysteresis even at $V_{\text{cell}} = 0.20$ V, although the hysteresis width relatively shrinks by the reduction of V_{cell}. The NI mode of the DM inverter at $V_{\text{cell}} = 1.2$ V shows almost the similar VTC to the conventional inverter except the logic threshold. The ST mode of the DM inverter at 0.20 V clearly exhibits the ST-inverter-type VTC. However, the VTC has much wider hysteresis than that of the ST inverter. This is due to the high pull-up ability of the pMOS FBT even at ultralow voltages.

Solid curves in Fig. 2 (c) shows butterfly curves of the 6T, ST, and ULVR cells for the ULV retention mode at $V_{\text{cell}} = 0.20$ V. The ULVR cell can considerably widen the operating-point-side lobe of the butterfly curve, which profits to enlarge the noise margin for the ULV retention. In addition, the ULVR cell has the wider lobe than the ST cell, which can be attributed to the difference in the current drivability of the FBTs at the ultralow voltage. Note that when the high level node is flipped to the other side node, the widened lobe is also moved to the flipped operating-point-side. Dotted curves in Fig. 2 (c) shows the butterfly curves of these cells for the worst case process corner (that is the SF or FS corner whose notations are defined later). Even in this case, the ULVR cell has a widespread lobe, i.e., a sufficient noise margin. More detailed analysis of the noise immunity is discussed later.

IV. DESIGN AND EVALUATION METHODOLOGIES

A. GENERAL PROCEDURE

Design of the ULVR cell with the HFPSs, hereafter referred to as the ULVR_{HF} cell, is carried out using HSPICE simulations with the careful consideration of variability of the constituent transistors. In this study, the LP model of the 65 nm silicon-on-thin-buried-oxide (SOTB) devices [39] is used for the CMOS transistors. The SOTB transistors enable efficient body bias control due to the extremely ultrathin buried-oxide layer between the substrate and the thin silicon channel layer.

Our developed design methodology is based on minimizing of the leakage power during the UVLR mode and maximizing/optimizing of the noise margin for all the operation modes (in particular, the READ and ULVR modes). Static noise margins (SNMs) are generally used for evaluation of stability in SRAM operations, where the opened loop circuitry for the cell is applied to the analysis. Nevertheless, for the ULVR cell, the analysis of noise immunity using SNMs is difficult (particularly in the ST mode), since the cell circuitry includes the internal feedback connections between the inverters and the delay for the feedback affects the noise hardness, i.e., the static analysis with the opened loop circuitry cannot precisely reproduce the internal feedback effects in the inverter loop. Therefore, dynamic noise margins (DNMs) with the closed loop circuitry shown in Fig. 3 (a) are used for the noise immunity analysis.

In the following analysis, various types of DNMs caused by noise sources placed at the N_1 and N_2 nodes in Fig. 3 (a) and at the $N_3 - N_5$ nodes in Fig. 1 (a) are investigated. The N_1- and N_2-node noises in the inverter loop are appropriate as indices for designing the ULVR_{HF} cell, and the N_3-, N_4-, and N_5-node noises on the power or bias lines are also used for evaluation of practical noise immunity of the cell. Note that introducing of the noise source at N_1 is qualitatively the same technique as the conventional SNM analysis. Noises

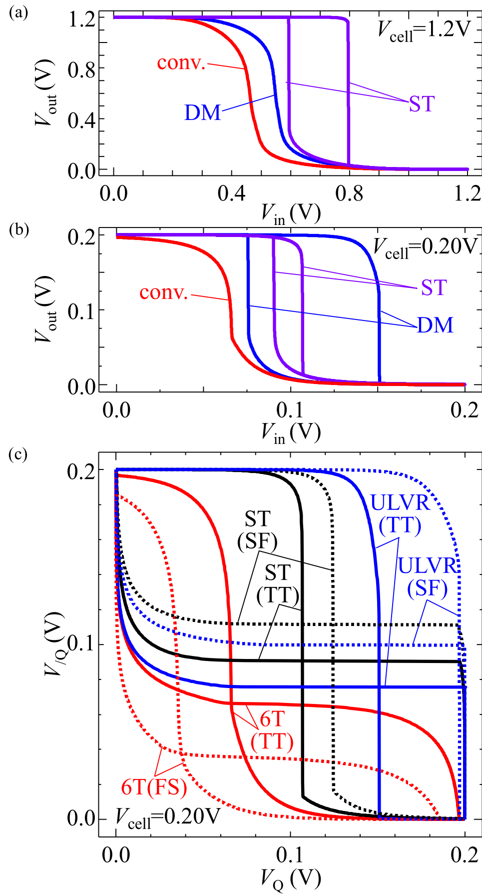

FIGURE 2. Voltage transfer characteristics of conventional, ST, and DM inverters at (a) V_{cell} = 1.2 V and (b) 0.20 V. (c) Butterfly curves of the 6T, ST, and ULVR cells at V_{cell} = 0.20 V. The doted curves show the butterfly curves for the worst-case corner.

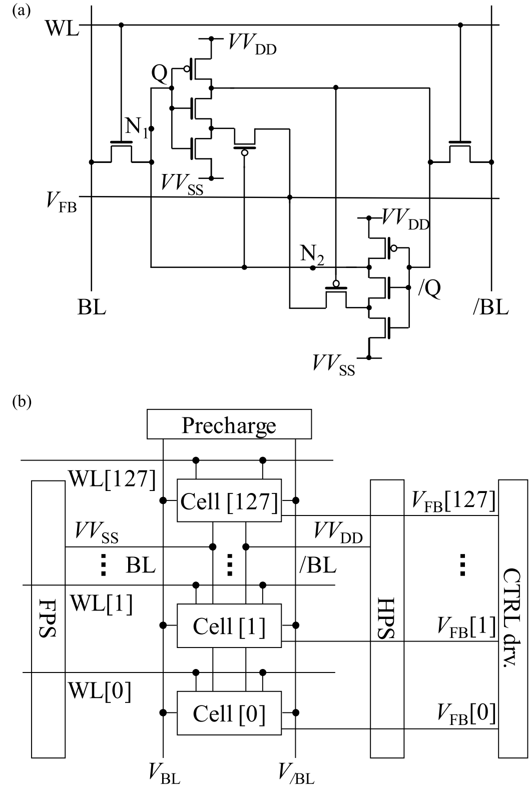

FIGURE 3. (a) Loop circuit representation of the ULVR-SRAM cell. (b) Array structure for analyzing cell characteristics and performances.

on power rails and bias lines could be more possible than the internal noises induced in the inverter loop. For the DNM analysis, rectangular-shape noise pulses with various pulse widths are used.

The circuit configuration shown in Fig. 3 (b) is used for the cell design and noise immunity analysis. A target cell for measuring DNMs with 127 dummy cells are connected with the bit lines BL and /BL, and the VV_{DD}, VV_{SS} and V_{FB} rails are also shared by these cells. The word lines WLs are individually connected to these cells. The CTRL driver is configured with the high threshold devices, which is shared by the 128 cells in the column. The sizes of the CTRL driver transistors are determined based on the latency of the mode switching. A channel width of 500 nm is used for both the pMOS and nMOS devices of the CTRL driver. The header power switches (PS_1 and PS_2) and the footer power switches (PS_3 and PS_4) are also comprised of the high threshold devices. PS_1 and PS_3 shared by the 128 column cells are designed so as to ensure sufficient noise margins for the normal SRAM operations. A channel width of 1000 nm is used for PS_1 and PS_3 for the 128 column cells. PS_2 and PS_4 are designed so as to satisfy a sufficient noise margin during the ULVR mode. A channel width of 100 nm per single cell is used for PS_2 and PS_4. The super cut-off (SC) technique [40] is applied for the shutoff of these power switches, where the gate voltages of 1.4 V and −0.20 V are employed for the pMOS and nMOS power switches, respectively. For the noise immunity analysis, Monte-Carlo simulations are carried out using a single cell with the substantively same peripherals for simplicity.

The bias conditions for all the operation modes are shown in Table 1. Note that the voltage of the bit lines during the ULVR mode is important to minimize the leakage power. In this study, the bit-line voltage during the ULVR mode is set to V_{SSH}. This bias induces the body bias effect to the pass transistors, effectively suppressing the leakage currents through the pass transistors.

B. CELL DESIGN

The DNM induced by the N_1-node noise in the ULVR mode is used as a design index for the $ULVR_{HF}$ cell. The N_1-node noise is more useful than the N_2-node noise, as shown later. Each DNM for the READ, SB (standby (retention) at V_{cell} = V_{cellH}), and ULVR modes is determined as a maximum noise voltage so that the holding data are not flipped by inducing a noise pulse. For the WRITE operation, the SNM for the N_1-node noise is used as an index in a traditional fashion (since the DNM analysis is difficult for this mode and the

TABLE 1. Operating conditions for the $ULVR_{HF}$ cell.

	V_{BL} (V)	V_{WL} (V)	V_{PS1} (V)	V_{PS2} (V)	V_{PS3} (V)	V_{PS4} (V)
ULVR	V_{SSH}	0.0	1.4	0.0	−0.20	1.2
READ	1.2	1.2	0.0	1.4	1.2	−0.20
WRITE	1.2 / 0.0	1.2	0.0	1.4	1.2	−0.20
SB	FL	0.0	0.0	1.4	1.2	−0.20

FL: Floating

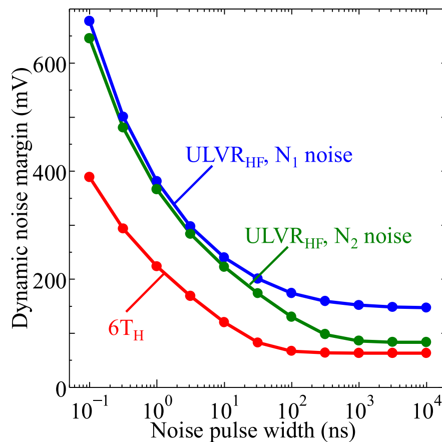

FIGURE 4. DNM as a function of noise pulse width for the $ULVR_{HF}$ and $6T_H$ cells during the ULVR mode.

SNM analysis is permissible for the NI mode (in particular, the WRITE operation).

Fig. 4 shows DNM as a function of noise pulse width for the $ULVR_{HF}$ cell during the ULVR mode, in which N_1- and N_2-node noises are applied to the $ULVR_{HF}$ cell. Here, the $ULVR_{HF}$ cell uses the optimum design shown in the next section, and the data in the figure are the results of the post-layout simulations. In this figure, DNM for the $6T_H$ cell (the subscript "H" represents the HPS architecture, defined later) is also shown, where the noise is applied on the inverter-loop interconnection of the cell (which is equivalent to the N_1-node noise). All the DNMs decreases with increasing pulse width, and these DNMs saturate for pulse widths longer than $\sim$1 μs ($ULVR_{HF}$ cell) or $\sim$0.1 μs ($6T_H$ cell). In the $ULVR_{HF}$ cell, the DNM for the N_2-node noise is lower than that for the N_1-node noise. In particular, this behavior much stands out for pulse widths longer than $\sim$1 μs. This is because the N_2-node noise directly weakens the feedback effect induced by the FBTs. Nevertheless, the DNM for the N_1-node noise with a sufficiently long pulse width (hereafter, it is referred to as a quasi-static noise margin (QSNM)) is appropriate as a design index, since it is helpful to maximize/optimize the feedback effect for noise immunity of the cell. Therefore, in the following cell design, the QSNM for the N_1-node noise is employed. For the following failure (yield) analysis, the N_2-node noise is employed with the N_1-node noise and the other rail noises. The noise pulse width of 10 μs is used for evaluating the QSNMs.

The effects of the global variation in the threshold voltages on the QSNMs are included in the design optimization using 3σ process corners. Hereafter, the notations XY = TT, FF, SS, FS, and SF are used for the process corners, where the T represents the typical threshold voltage and F and S represent the 3σ-lower and -higher threshold voltages (in this study, the σ value of 20 mV is used for the global variation [41]). Thus, the T, F, and S devices represent the typical, fast, and slow devices, respectively. The process corner XY denotes the set of nMOS with the X-corner threshold voltage and pMOS with the Y-corner threshold voltage.

The design of the $ULVR_{HF}$ cell is optimized as follows: Firstly, a tentatively optimum design of the cell is determined using a pre-layout netlist. The optimization is carried out for maximizing the retention stability (QSNM), minimizing the leakage power, and satisfying the failure probability (yield) during the ULVR mode. Then, the cell is laid out based on this tentative design and parasitic elements are extracted from the layout. Subsequently, the cell is re-optimized using the post-layout netlist with these parasitic elements. Finally, the cell characteristics/performances are analyzed using the final post-layout netlist of the optimized cell design.

An 8KB ULVR-SRAM macro, which consists of the cell array and peripheral circuits including the PSs and CTRL drivers, is designed for analyzing the standby power and energy performances. Post-layout large-scale simulations are performed for the designed macro using FineSim. In this study, the ULVR-SRAM macro is analyzed assuming that various voltages and control signals are supplied from external power sources. For future system applications, it is necessary to evaluate the overheads in terms of circuit area, power, and cost required to generate these voltages and signals. Note that recent MPs and SoCs generate a large number of different voltages using an on-chip DC-DC converter [42] that can be applied to the ULVR-SRAM.

C. NOISE IMMUNITY AND FAILURE ANALYSIS

The noise immunity and resulting failure probability of the $ULVR_{HF}$ cell are analyzed by Monte-Carlo simulations, where the random local variation of the devices is carefully considered for the QSNM analysis. Although, in this case, the characteristics of the constituent devices vary individually, the probability of occurrence can be treated statistically. The N_1- and N_2-node noises (see Fig. 3 (a)) are used for the failure probability analysis. The worst-case QSNM in the ULVR mode depends on not only the retention state (given by V_Q and $V_{/Q}$) but also the position of the noise source (note that there are the N_1- and N_2- equivalent /Q-side positions in the inverter loop). Therefore, four patterns depending on the configurations of retention state ((V_Q, $V_{/Q}$) = (H, L) or (L, H)) and noise-source position (Q-side or /Q-side) are examined for N_1- and N_2-node-noise-induced QSNMs. A pulse width of 10 μs is used for the failure probability analysis. The normal probability distribution is used for the random local variation in the threshold voltages of the pMOS and nMOS devices. The channel-area-dependence of

the local variation distribution of the threshold voltage in each transistor is formulated using the following standard deviation: $\sigma_L = \sigma_{L0}\ (L_0 W_0\ /\ LW)^{1/2}$, in which L_0 and W_0 represent the minimum design sizes of channel length and width, respectively, and σ_{L0} is the standard deviation for the minimum device design. In this study, the following parameters are used: $L_0 = 60$ nm $W_0 = 100$ nm, and $\sigma_0 =$ 20.0 mV for the pMOS devices and 21.8 mV for the nMOS devices [43]. The number of trials for each Monte-Carlo simulation is set to 100000. The N_3-, N_4-, and N_5-node noises (see Fig. 1) on the WL, VV_{DD}, and V_{FB} rails are also used for noise-immunity analysis of the cell.

The noise immunity can be qualitatively evaluated by the position of noise-induced QSNM distribution obtained from the above-described Monte-Carlo simulation. The allowance of the worst-case tail of noise-induced QSNM distribution can be analyzed using the cumulative distribution function $f_{CDF}(m)$ of the distribution. $f_{CDF}(m)$ represents the probability of the failure cells, where m is a QSNM value. In this study, 6σ failure probability (P_{FC}) is used for the probe of the failure cells. The allowable tail edge of QSNM distribution is given by m_0 that satisfy $f_{CDF}(m_0) = P_{FC}$. When m_0 is positive (or greater than a positive small number such as $\sim kT/q$), the cell array can be realized with the 6σ failure probability for the constituent cells. Note that the 6σ failure probability gives yield values of 99.9, 99.8, and 98.4% for 64KB, 256KB, and 2MB cell arrays, respectively.

D. COMPARATIVE STUDY

The characteristics and performances of $ULVR_{HF}$ cell are compared with those of a conventional 6T cell using the HPS architecture. In addition, from the discussions described in Section II, 6T, iso-area 6T (that has the same cell area as the $ULVR_{HF}$ cell), and ST cells, which have HFPSs, are also considered for the comparative study. Hereafter, these cells are simply referred to as $6T_H$, $6T_{HF}$, i-$6T_{HF}$, and ST_{HF} cells, respectively, where H and HF represent the HPS and HFPS architectures, respectively. When the PS architectures need not be specified, these suffixes are omitted.

The transistor sizes of the $6T_H$ cell are determined by reference to a 65 nm bulk SRAM design [44]. The channel widths W_{DRV}, W_{LD}, and W_{pass} of the driver, load, and pass transistors in the $6T_H$ cell are set to $W_{DRV} = 150$ nm, $W_{LD} = 100$ nm, and $W_{pass} = 100$ nm. The $6T_{HF}$ cell uses the same design as the $6T_H$ cell. For the i-$6T_{HF}$ cell, the channel width of each transistor in the cell is enlarged so that the cell area is almost identical with that of the $ULVR_{HF}$ cell, where the ratio of $W_{LD} : W_{DRV} : W_{pass}$ is kept constant as that of the $6T_H$ cell. This resizing method using only the channel widths is often used to evaluate an iso-area 6T cell for a ST cell [36], [37]. Fig. 1 (b) shows the circuit configuration of the ST_{HF} cell with the HFPSs. The ST_{HF} cell is designed so as to maximize its QSNM for the ULVR mode in the same manner as the $ULVR_{HF}$ cell (discussed in Section V).

TABLE 2. Design results for the reference SRAM cells.

	W_{LD} (nm)	W_{DRV} (nm)	W_{pass} (nm)	W_{FB} (nm)	V_{DDL} (V)	V_{SSH} (V)
$6T_H$	100	150	100	-	0.20	-
$6T_{HF}$	100	150	100	-	0.90	0.70
i-$6T_{HF}$	200	300	200	-	0.80	0.60
ST_{HF}	100	110	100	190	0.50	0.30

V_{DDH}=1.2V, V_{SSL}=0.0V

The design results of these reference cells are shown in Table 2. The $6T_{HF}$, i-$6T_{HF}$, and ST_{HF} cells use the same HFPS design as the $ULVR_{HF}$ cell. The $6T_H$ cell use the same header PS design as that in the HFPSs of the $ULVR_{HF}$ cell. The channel-area-dependent local-variation distribution of the threshold voltage in each transistor is determined from the relation of $\sigma_L \propto 1/(LW)^{1/2}$, as described above.

V. DESIGN OPTIMIZATION AND CIRCUIT PERFORMANCE

A. DESIGN OPTIMIZATION

For the following cell design, the channel length and width of the load and pass transistors are fixed at 60 and 100 nm, respectively (that are the minimum channel length and width of the pMOS and nMOS devices). The channel widths W_{DRV} and W_{FB} of the driver transistors and FBTs, respectively, are optimized for various V_{SSH} values with a constant V_{cell} (= V_{cellL}) condition in the ULVR mode, which have a great impact on the stability and leakage during the ULVR mode. The target value of the lowest QSNM due to the global variation is set to 100 mV that can ensure a sufficiently low failure probability discussed later. Namely, W_{DRV} and W_{FB} are optimized under the constraint of QSNM $\geq$ 100 mV so that the leakage power can be minimized.

Figs. 5 (a)–(d) show the contour mapping of QSNM for the $ULVR_{HF}$ cell, where V_{SSH} is varied from 0.20 V to 0.50 V with the condition of $V_{cellL} = V_{DDL} - V_{SSH} =$ 0.20 V. The vertical and horizontal axes are W_{DRV} and W_{FB}, respectively. All the plot points in each figure represent the worst-case QSNM values due to the global variation under the given conditions of V_{SSH}, W_{DRV}, and W_{FB}, and the brightest regions represent the allowable QSNM condition (QSNM $\geq$ 100 mV). The brightest (allowable QSNM) regions appear depending on V_{SSH}, which results from the balance of the body bias effects of the pMOS and nMOS transistors. Lower V_{SSH} conditions weaken the body bias effect of the nMOS transistors and enhance that of the pMOS transistors, and higher V_{SSH} conditions cause these opposite results. As a result, the allowable QSNM region is maximally widened at $V_{SSH} = 0.30$ V ($V_{DDL} = 0.50$ V). The region extends to the lower left, resulting in a small cell size.

Fig. 6 shows variations of QSNM and leakage power along the line A-B on the allowable QSNM region shown in Fig. 5 (c). The SF-corner margin is sensitive to W_{DRV} and it determines the minimum QSNM for $W_{DRV} \leq 150$ nm. On the other hand, the FS-corner margin is independent of W_{DRV}, which governs the minimum QSNM for $W_{DRV} >$

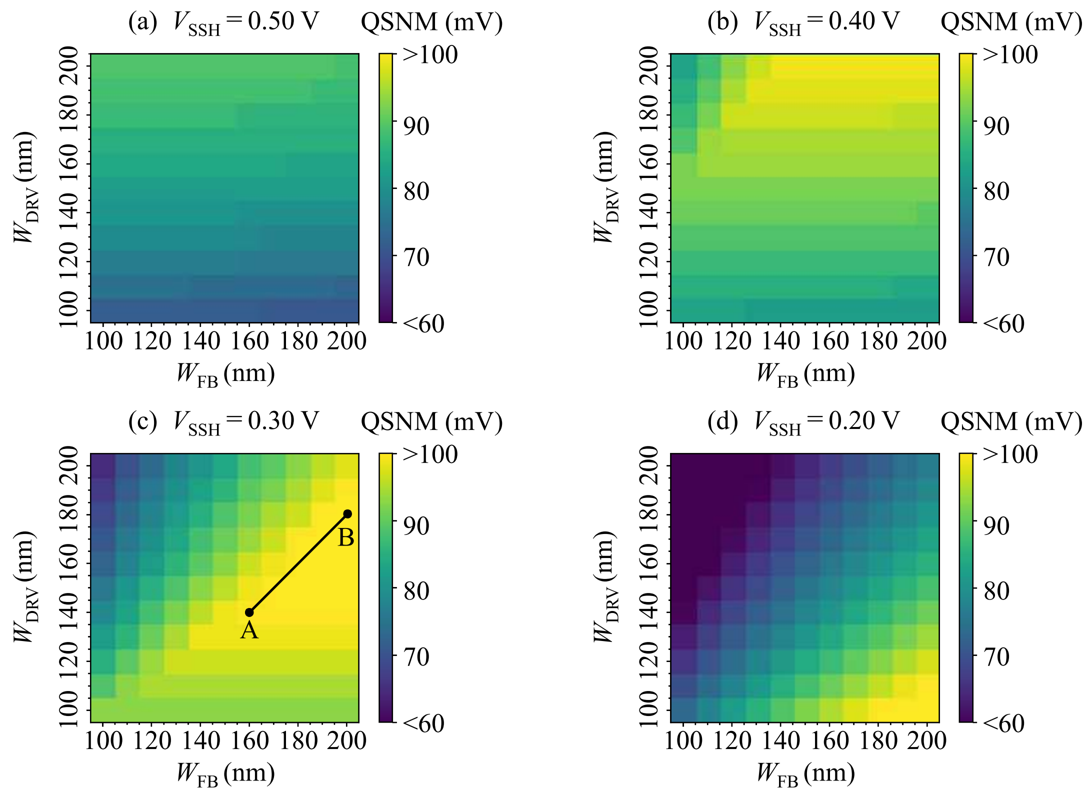

FIGURE 5. Contour mapping of QSNM for the $ULVR_{HF}$ cell during the ULVR mode. V_{DDL} is given by V_{DDL} = 0.20 V + V_{SSH}.

TABLE 3. Optimized design parameters of the $ULVR_{HF}$ cell.

L	60 nm	W_{pass}	100 nm
W_{LD}	100 nm	V_{DDH} / V_{DDL}	1.2 V / 0.50 V
W_{DRV}	140 nm	V_{SSH} / V_{SSL}	0.30 V / 0.0 V
W_{FB}	160 nm		

150 nm. The leakage power monotonously decreases with decreasing W_{DRV} and W_{FB} along the line A-B on Fig. 5 (c). The leakage power is minimized at the lower left edge of the allowable QSNM region. Therefore, a design of W_{DRV} = 140 nm, W_{FB} = 160 nm, and V_{SSH} = 0.30 V with V_{cellL} = 0.20 V is used for the cell. The detail of the cell design is summarized in Table 3.

The design methodology shown here can be applied for other design rules based on different devices/processes. Although the retention voltage (V_{cellL}) needs to optimize for each design rule, this can be easily accomplished using the same method shown in Fig. 5 with various V_{cellL} values. The results can be checked by the failure analysis shown later.

Figs. 7 (a) and (b) show cell layouts of the control $6T_H$ cell and the optimally designed $ULVR_{HF}$ cell, respectively. Although the $ULVR_{HF}$ cell has a ST-inverter-based 10T structure, the cell area is relatively suppressed compared with conventional 10T ST-inverter-based cells [36], [37]. This is because the pMOS FBTs enable compactly to arrange all the transistors in the cell, i.e., the FBTs can be placed in the same well with the load transistors. The area penalty of the $ULVR_{HF}$ cell is 77% in comparison with the $6T_H$ cell. Fig. 7 (c) shows a cell layout of the optimally designed ST_{HF} cell. Although the dead space accrues in the layout of conventional 10T ST cells owing to their nMOS FBTs [36], the layout shown in Fig. 7 (c) can reduce the dead space. The dead space can be eliminated by the adjacent stepwise-edge-shape cells. This layout can be achieved by changing the arrangement of the driver and pass transistors from the conventional layout. The area penalty of the ST_{HF} cell is 87% in comparison with the $6T_H$ cell.

Note that the scalability of the $ULVR_{HF}$ cell is considered qualitatively equivalent to that of ST-based cells. For these ST cells, the scaling is effective at improving the noise margin [37]. In particular, advanced devices such as FinFETs, which have higher current drivability and lower off-current ability, could be beneficial for enhancing the performance of the $ULVR_{HF}$ cell. Nevertheless, when advanced devices are used, the cell design/size and array area efficiency need to be re-examined.

Fig. 8 (a) shows the QSNMs of the $ULVR_{HF}$ cell for the ULVR, READ, and SB operations and the SNM for the WRITE operation, which are obtained from the post-layout analysis. The worst case is the SF corner for the ULVR operation, whose margin exceeds 100 mV. Note that this margin value for the 3σ-corners is required to satisfy a sufficiently low failure probability, discussed later. The worst-case READ margin (FS corner) is also close to this value. However, the READ margin can be easily gained by introducing bias techniques such as word-line

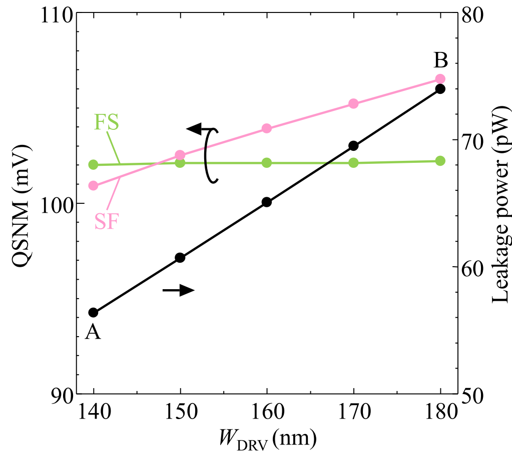

FIGURE 6. Variations of SF- and FS-corner QSNMs and leakage power along the line A-B on the allowable QSNM region shown in Fig. 5 (c).

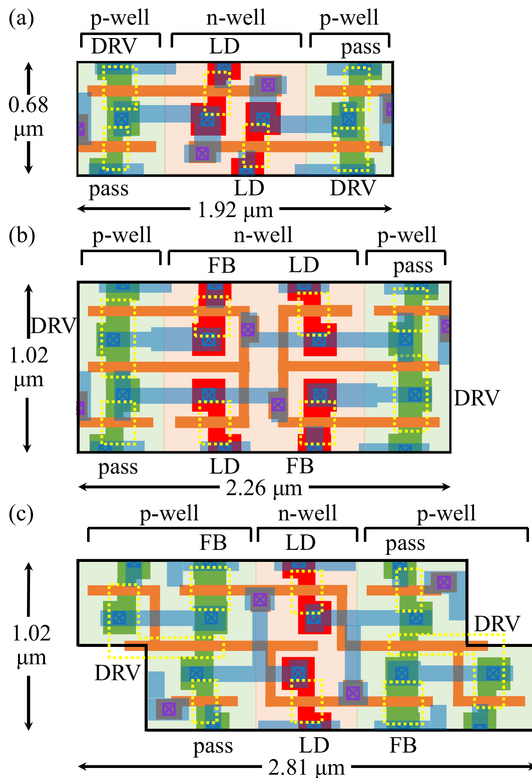

FIGURE 7. Layouts of the designed (a) $6T_H$, (b) $ULVR_{HF}$, and (c) ST_{HF} cells.

underdrive [45]. Therefore, the ULVR mode is employed for the following noise immunity and failure analyses.

Fig. 8 (b) shows the QSNMs of the $ULVR_{HF}$ cell and the other reference cells for the ULVR mode. The QSNMs of the $6T_{HF}$ and i-$6T_{HF}$ cells are enhanced compared with the $6T_H$ cell. However, the degree of enhancement is not as high as that of the ST_{HF} or $ULVR_{HF}$ cell. The $ULVR_{HF}$ cell has the higher QSNMs than the ST_{HF} cell. This is due to the difference in the FBTs, i.e., the pMOS FBTs are more effective.

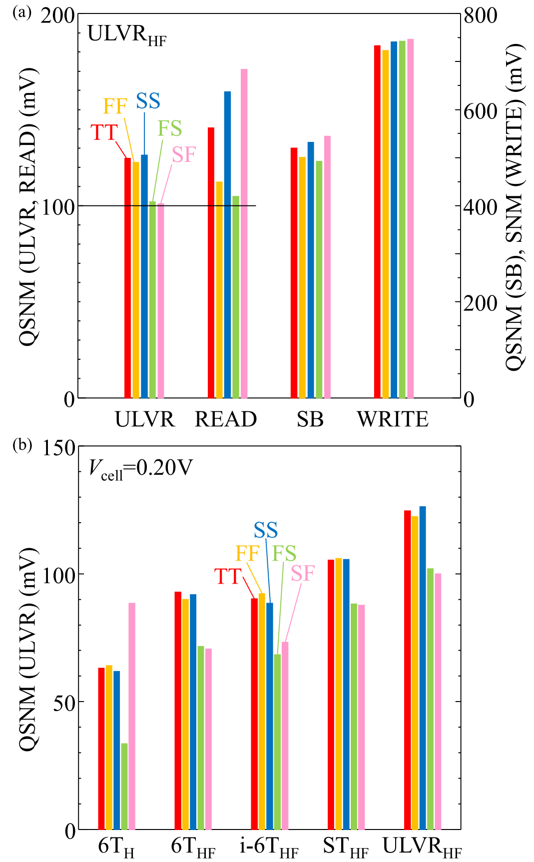

FIGURE 8. (a) QSNMs of the optimally designed $ULVR_{HF}$ cell for the ULVR, READ, and SB operations and the SNM for the WRITE operation (b) QSNMs of the $ULVR_{HF}$ cell and the other reference cells for the ULVR operation.

Fig. 9 shows the Monte-Carlo simulation histogram of QSNMs for the $ULVR_{HF}$ cell and the other reference cells in the ULVR mode, in which the N_1-node noise is examined under the random local variation in threshold-voltage for the constituent transistors in the cells. The effect of the channel area on the local variation is included in these simulations (see Section IV). The ULVR proceeds at V_{cell} = 0.20 V. For the 6T cells, changing from the HPS to the HFPSs improves the noise immunity. In addition, the i-$6T_{HF}$ cell can reduce the dispersion. However, these cells show weaker noise immunity than the ST_{HF} and $ULVR_{HF}$ cells. The distribution of the $ULVR_{HF}$ cell is further shifted to the higher QSNM region.

The allowance of the worst-case tail of N_1-node-noise-induced QSNM distributions can be analyzed using the relation of $f_{CDF}(m_0) = P_{FC}$ (see Section IV). A positive m_0 value (or greater than a positive number such as $\sim kT/q$) can realize the cell array with the 6σ failure probability for the constituent cells (i.e., with feasible redundancy). The cumulative distribution function (CDF) of the QSNM distributions for the N_1-node noise is shown by solid circles in Fig. 10, where the data are fitted by the superposition of several Gaussians. The $ULVR_{HF}$, ST_{HF}, and i-$6T_{HF}$ cells have satisfactorily positive QSNM edge (m_0) values ($> \sim kT/q$), i.e., these cells satisfy the 6σ failure probability. Note that, as

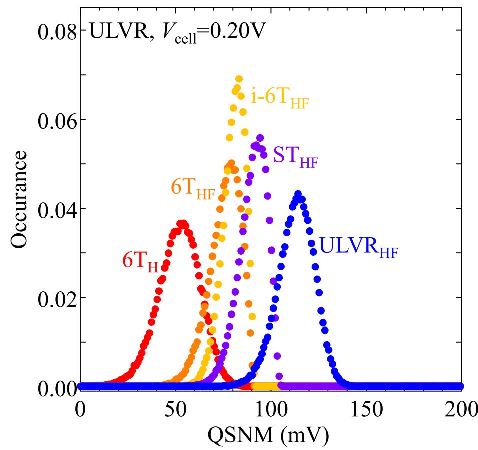

FIGURE 9. Monte-Carlo simulation histogram of QSNMs for the ULVR$_{HF}$ cell and the other reference cells in the ULVR mode.

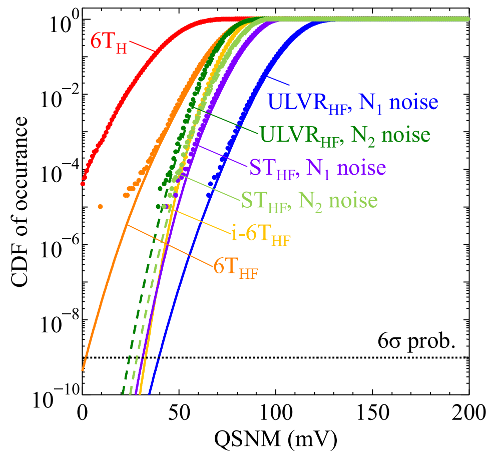

FIGURE 10. Cumulative distribution function of the QSNM distributions for the ULVR$_{HF}$ cell and the other reference cells in the ULVR mode. The dotted line represents the 6σ failure probability.

shown by the dashed curves in the figure, both the ULVR$_{HF}$ and ST$_{HF}$ cells show sufficient noise immunity (m_0 = $\sim kT/q$) even for the N$_2$-node noise, although the QSNM for the N$_1$-node noise is used as a design index for optimizing the transistor sizes. This means that the QSNM for the N$_1$-node noise can be satisfactorily used as a design index. Although employing the QSNM for the N$_2$-node noise as an additional design index can refine the cell design, this loses the simplicity.

B. NOISE IMMUNITY

Fig. 11 shows the CDFs of the ULVR$_{HF}$ cell and the other reference cells for the N$_3$-node noise (that represents the VV_{DD} rail noise) during the ULVR mode at V_{cell} = 0.20 V. The CDF tails of the ULVR$_{HF}$, ST$_{HF}$, and i-6T$_{HF}$ cells can satisfy the 6σ failure probability (i.e., positive m_0) for the 0.2V-retention mode. On the other hand, the 6T$_{H}$ and 6T$_{HF}$ cells cannot satisfy this criterion.

Table 4 summarizes the results (m_0 values) of failure analysis for all the node noises during the ULVR mode at V_{cell} = 0.20 V. The ULVR$_{HF}$ cell shows the satisfactorily positive m_0 values ($> \sim kT/q$) for all the node noises. The ST$_{HF}$ and i-6T$_{HF}$ cells also satisfies the criterion for the 6σ failure probability. On the other hand, the 6T$_{H}$ and 6T$_{HF}$ cells cannot satisfy the criterion. Note that for the 6T$_{HF}$ cell, the criterion (positive m_0 value) can be satisfied by relieving the failure probability to 5.5σ, when V_{cell} is enlarged to 0.65 V (with V_{SSH} = 0.50 V) and the pulse width is reduced to less than 10 μs.

TABLE 4. Failure analysis results (m_0 values) for the all the node noises in the ULVR mode at V_{cell} = 0.20 V.

	N$_1$	N$_2$	N$_3$	N$_4$	N$_5$
6T$_{H}$	Ng	-	Ng	Ng	-
6T$_{HF}$	1.80 mV	-	Ng	726 mV	-
i-6T$_{HF}$	32.7 mV	-	40.5 mV	684 mV	-
ST$_{HF}$	31.2 mV	28.3 mV	27.5 mV	243 mV	154 mV
ULVR$_{HF}$	39.6 mV	24.3 mV	38.7 mV	253 mV	82.4 mV

Ng: Negative value

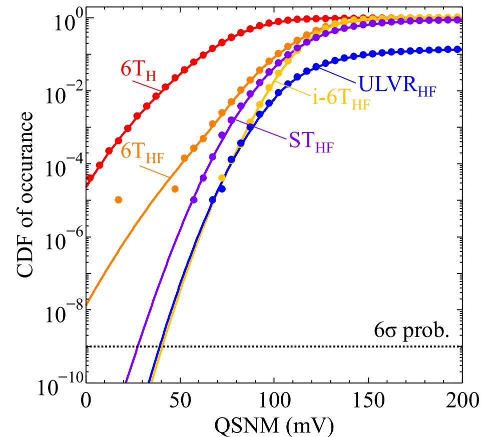

FIGURE 11. Cumulative distribution function of the N$_3$-node-noise-induced QSNM distributions for the ULVR$_{HF}$ cell and the other reference cells in the ULVR mode.

C. POWER PERFORMANCE

Fig. 12 shows leakage power for the ULVR$_{HF}$ cell and the other reference cells during the retention operations. The framed bars represent the leakage power of each cell in the SB mode at V_{cell} = 1.2 V. The filled bars represent the leakage power in the following operating states: Low-voltage retention (LVR) with V_{cell} = 0.65 V for the 6T$_{H}$ and 6T$_{HF}$ cells, and ULVR with V_{cell} = 0.20 V for the ULVR$_{HF}$, ST$_{HF}$, and i-6T$_{HF}$ cells. Note that the 6T$_{H}$ and 6T$_{HF}$ cells cannot satisfy the same degree of noise immunity as the ULVR$_{HF}$ cell even for V_{cell} = 0.65 V, as noted above. The condition of V_{cell} = 0.65 V is used as a reference for 6T$_{H}$ and 6T$_{HF}$ cells. The V_{SSH} values for the ULVR$_{HF}$ and reference cells in the ULVR/LVR mode are shown in Tables 2 and 3, which are determined by optimizing the noise immunity and leakage power of each cell, as discussed previously. In this figure, a NV-SRAM cell (NV$_{H}$ cell) using magnetic tunnel junctions (MTJs) with the HPS is also examined [19], where

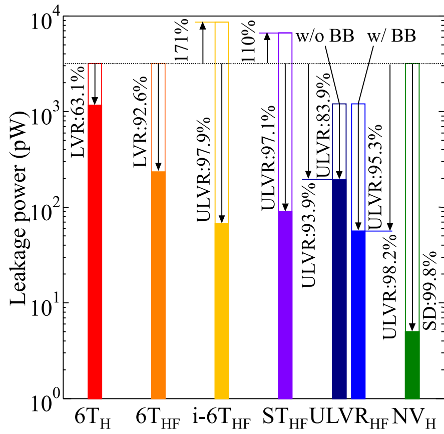

FIGURE 12. Leakage power of the $ULVR_{HF}$ cell and the other reference cells during the various retention operations. The cases with and without the body bias (BB) effect are examined for the $ULVR_{HF}$ cell.

the complete power shutdown (SD) mode is applied to the cell. The bistable circuit part of the NV_H cell uses the same design as the $6T_H$ cell.

The analysis use the circuit system shown in Fig. 3 (b). The bit lines for all the HFPS-architecture cells are set to the floating state during the SB mode at $V_{cell} = 1.2$ V and to V_{SSH} for the ULVR/LVR mode. The $6T_H$ cell uses the bit line states of floating and V_{DDL} for the 1.2V- and 0.65V-retention modes, respectively. For all the HFPS-architecture cells, the body biases to the constituent transistors are automatically applied during the ULVR/LVR mode using the body-bias method described in Section III. The case without the body bias effect is also examined for the $ULVR_{HF}$ cell, which is achieved by connecting the body terminals to the VV_{DD} or VV_{SS} rail.

For the $6T_H$ cell, changing the retention mode from SB (at $V_{cell} = 1.2$ V) to LVR (at $V_{cell} = 0.65$V) can reduce the leakage power by 63.1%. It can be improved to 92.6% by introducing the HFPS architecture (i.e., the $6T_{HF}$ cell), which is due to the voltage dividing effect of the HFPSs and the body bias effect during the LVR mode. For the i-$6T_{HF}$ cell, although the leakage power can be further reduced using the ULVR at $V_{cell} = 0.20$ V, it significantly increases during the SB mode owing to the increase in the channel region. The ST_{HF} cell can also effectively reduce the leakage power during the ULVR mode. Nevertheless, the leakage power does increase during the SB mode at $V_{cell} = 1.2$ V, owing to the high leakage current through the FBTs.

The $ULVR_{HF}$ cell can dramatically reduce the leakage power using the ULVR at $V_{cell} = 0.20$ V. The reduction rate from the standby power of the $6T_H$ cell attains to 98.2%. The body biases induced automatically only during the ULVR mode are effective at reducing the leakage power, as shown in the figure. The leakage-power reduction ability of the ULVR mode is almost comparable to that of the SD mode of the NV_H cell, i.e., the ULVR is efficient and the effect is close to the nonvolatile retention. Note that the leakage power of the $6T_H$ cell in the LVR mode is significantly (∼20 times) higher than that of the $ULVR_{HF}$ cell in the ULVR mode.

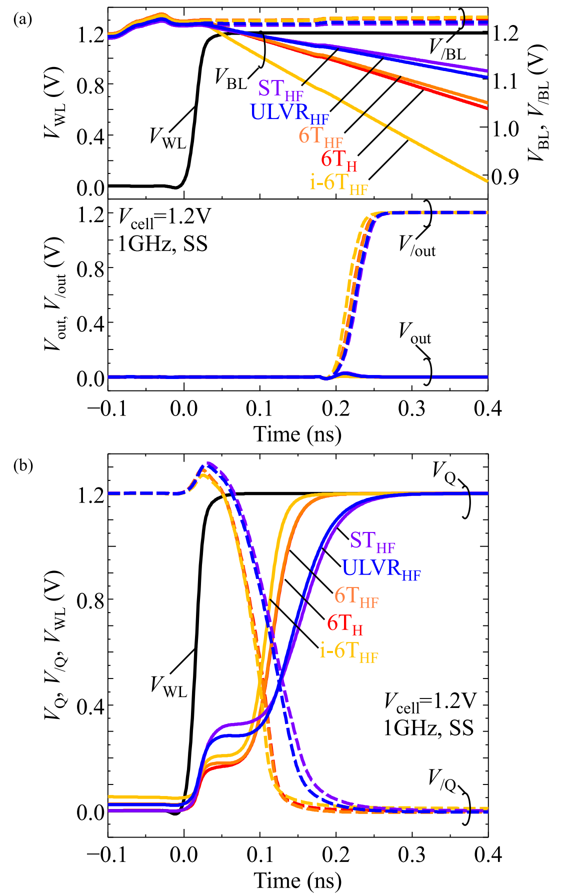

FIGURE 13. (a) Simulated waveforms of the bit-line voltages (V_{BL} and $V_{/BL}$), word line voltage (V_{WL}), and output voltages (V_{out} and $V_{/out}$) during the READ operation at $V_{cell} = V_{cellH}$. (b) Simulated waveforms of the storage-node voltages (V_Q and $V_{/Q}$) and V_{WL} during the WRITE operation at $V_{cell} = V_{cellH}$.

The $ULVR_{HF}$ cell can also reduce the leakage power during the SB mode at $V_{cell} = 1.2$ V. This is due to the effects of the stacking driver transistors and of the V_{FB} bias. Note that V_{FB} for the $ULVR_{HF}$ cell is set to 0.0 V in the NI mode. On contrary, V_{FB} for the ST_{HF} cell is always set to VV_{DD} (see Fig. 1 (b)).

D. READ AND WRITE OPERATIONS

The upper panel in Fig. 13 (a) shows the transient response of the bit line voltages V_{BL} and $V_{/BL}$ during the READ operation of the $ULVR_{HF}$ cell and the other reference cells for $(V_Q, V_{/Q}) = (L, H)$, where the NI mode is used for the $ULVR_{HF}$ cell. The evaluation (peripheral) circuit is the same as the circuit configuration of the ULVR-SRAM macro (shown later, see Fig. 14 (a)) except the decoder. The clock frequency is 1 GHz. The worst-case process corner SS is employed for the transistors in each cell. The $ULVR_{HF}$ and ST_{HF} cells have the almost same discharge speed, which cause ∼40% degradation from the $6T_H$ cell. Note that the $6T_{HF}$ cell has the almost identical discharge speed with the

6T$_H$ cell, meaning that the HFPSs used in this study has no adverse influence on the discharge speed. The i-6T$_{HF}$ cell shows the fastest discharge speed. These results can be attributed to the difference in the size of the driver transistors in each cell. The lower panel in Fig. 13 (a) shows the output signals V_{out} and $V_{/out}$ of the sense amplifier. Although the discharge speed of the ULVR$_{HF}$ cell is slower than that of the 6T cells, the responses of the output signals from the sense amplifier are almost the same for all the cells. The ULVR$_{HF}$ cell can achieve the READ operation at several GHz.

Fig. 13 (b) shows the transient response of the storage node voltages V_Q and $V_{/Q}$ during the WRITE operation for all the cells, which shows the case that the condition of (V_Q, $V_{/Q}$) = (L, H) is flipped to (H, L). As with Fig. 13 (a), the worst-case process corner SS is employed for all the cells, and the ULVR$_{HF}$ cell operates as the NI mode. Although the flipping speeds are varied depending on the types of cells, the write operation is almost completed within ~200 ps, which allows to operate at several GHz. From the results shown in Figs. 13 (a) and (b), it can be concluded that the ULVR$_{HF}$ cell with the NI mode has the comparable performances to the 6T cells for the normal SRAM operations.

It is worthy to note that using the ST mode, the ULVR$_{HF}$ cell can perform low-voltage SRAM operations even at the minimum-energy point V_{Emin} (~0.4 V), which can dramatically reduce the active power [34], [46]. The cell with the ST mode can have satisfactorily noise immunity required for the READ and WRITE operations at V_{Emin}. The worst-case READ margin sufficiently satisfies the 6σ failure probability, although the cell has no assist circuits for low-voltage SRAM operations described in Section II. This noise immunity might be inferior to cases using assist circuits. Nevertheless, the ST mode of the ULVR$_{HF}$ cell is also effective at achieving the noise immunity required for the $V_{Emin.}$ operations. Therefore, the ULVR$_{HF}$ cell can achieve the high performance SRAM operations at V_{cellH}, active-power-reduced SRAM operations at V_{Emin} ($V_{cellL} < V_{Emin} < V_{cellH}$), and significant standby-power reduction using the ULVR mode at V_{cellL}.

E. DESIGN AND CHARACTERIZATION OF ULVR-SRAM MACRO

Finally, power and energy performances of the ULVR$_{HF}$-SRAM are analyzed using its 8KB macro. Figs. 14 (a) and (b) show the block diagram and layout of the 8kB ULVR$_{HF}$-SRAM macro, respectively, where the cell design and layout described above are employed. The HFPSs are arranged on the left and right sides of the subarray blocks, as shown in Fig. 14. The VV_{DD} and VV_{SS} rails extend in the bit-line direction on the cell array region, and these are connected to the HFPSs using the interconnections in the word-line direction. The M4 and M5 layers are used for these rails and interconnections, respectively. The V_{FB} lines extend in the word-line direction using the M3 layer. All these additional rails can fit inside the pitch of the cell

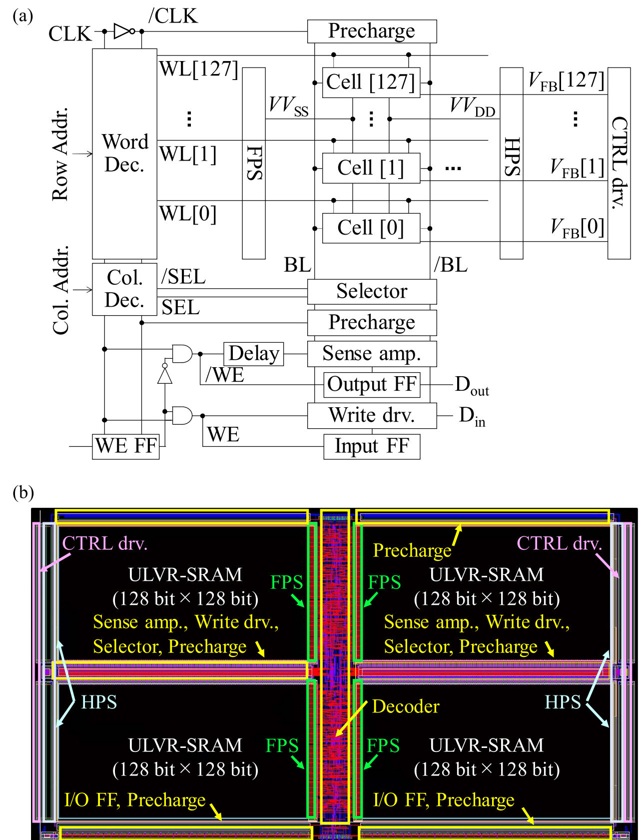

FIGURE 14. (a) Block diagram and (b) layout of the 8KB ULVR$_{HF}$-SRAM macro.

layout, and the layout of the cell array can be arranged without any dead space, i.e., the various types of rails do not affect the array area efficiency. The area overheads of the HFPSs and CTRL drivers are 4.8 and 1.1% of the total area of the macro, respectively. An 8KB macro of the reference 6T$_H$-SRAM is also designed. These designed macros are evaluated by post-layout large-scale simulations using FineSim.

Fig. 15 shows the leakage power of the ULVR$_{HF}$-SRAM macro, where the SB and ULVR modes are applied to the cell array and the SB and SD modes to the peripherals, i.e., the conditions of (cell array, peripheral) = (SB, SB), (SB, SD), and (ULVR, SD) are examined. The results of the 6T-SRAM macro are also shown in the figure, where the LVR is carried out at V_{cell} = 0.65 V. The ULVR mode of the ULVR$_{HF}$-SRAM macro can reduce the standby power by 97.6% from that in its SB mode. The standby power reduction can reach to 98.5% compared with the SB mode of the 6T$_H$-SRAM macro. Thus, the PG operation using the ULVR mode is highly effective for the cell array. The LVR mode of the 6T$_H$-SRAM macro can reduce the leakage power by 61.7%. These results are almost identical with those of the cell-level evaluations shown in Fig. 12. The maximum operating frequency of the ULVR$_{HF}$-SRAM macro is 600 MHz. As described in the previous section, the read/write operations at 1 GHz or more are possible for the cell. Although the macro

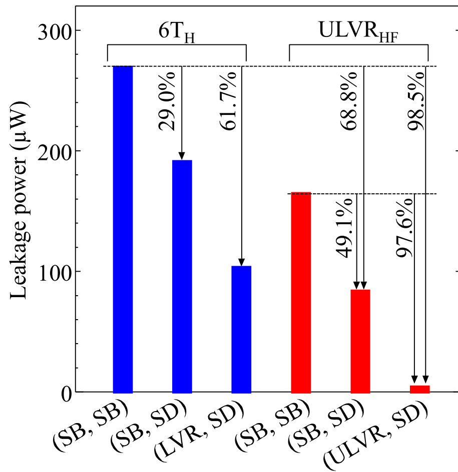

FIGURE 15. Leakage power of the 8KB $ULVR_{HF}$- and $6T_{H}$-SRAM macros. The operating conditions of (cell array, peripheral) = (SB, SB), (SB, SD), and (ULVR/LVR, SD) are examined.

uses exactly the same cell design, the design of its decoder is not optimized. The operating frequency is restricted to 600 MHz by the decoder.

The PG granularity of the $ULVR_{HF}$-SRAM is analyzed using a performance index BET that is defined as a minimum ULVR duration to compensate for the excess energies E_{ENT} and E_{EXT} caused by the enter (ENT) operation to the ULVR mode and the exit (EXT) operation from the ULVR mode, respectively. The BET (T_{BE}) is given by $T_{BE} = (E_{ENT} + E_{EXT}) / (P_{SB} - P_{ULVR})$, where P_{SB} and P_{ULVR} represent the leakage power during the SB and ULVR modes, respectively. E_{ENT} and E_{EXT} depend on the speed for the mode transition (the ramp rate for the power switch control). In this study, the ramp rates for the ENT and EXT operations are determined from the allowable current of the interconnects. The parameters for the BET calculation are shown in Table 5, which are obtained by analyzing the 8KB macro using FineSim. T_{BE} is calculated to 1.51 μs for the 8KB macro. From this result, the BET values for 32KB, 256KB and 2MB ULVR-SRAMs are estimated to 1.55, 1.94, and 5.00 μs, respectively, that are much lower than BET values ($\sim$100 μs to several ms) for NV-SRAMs using the nonvolatile retention [19]. These lower BET values of the $ULVR_{HF}$-SRAMs would enable them to achieve fine-grained PG using the ULVR.

Finally, perspective on the ULVR-SRAM technology is briefly illustrated. The $ULVR_{HF}$ cell has the advantage of smaller cell size than the ST_{HF} cell, as shown in Fig. 7. However, the size is still much larger than that of conventional 6T cells. Nevertheless, the $ULVR_{HF}$ cell has the extremely high reduction ability of leakage power using the stable ULVR operation, which is inaccessible to conventional 6T cells, as shown in Figs. 8–12. In addition, the energy overhead for the PG operation using the ULVR is sufficiently low, i.e., the $ULVR_{HF}$ cell can conduce to lower BETs, as shown in Table 5, which are much lower values than the cases using nonvolatile retention. Moreover, the

TABLE 5. BET values of $ULVR_{HF}$-SRAMs and parameters using these calculations.

P_{SB}	164 μW	t_{EE}	26.1 ns
P_{ULVR}	3.88 μW	T_{BE} (8KB)	1.51 μs
E_{EE}	243 pJ	T_{BE} (32KB)	1.55 μs
τ_{ENT}	12.0 ns	T_{BE} (256KB)	1.94 μs
τ_{EXT}	14.1 ns	T_{BE} (2MB)	5.00 μs

t_{ENT} (t_{EXT}): Total latency required for turn-off of PS_1 and PS_3 (PS_2 and PS_4) and successively for turn-on of PS_2 and PS_4 (PS_1 and PS_3).
$E_{EE} = E_{ENT} + E_{EXT}$, $t_{EE} = t_{ENT} + t_{EXT}$

$ULVR_{HF}$ cell enables the high-performance SRAM operations at the ordinary supply voltage, as shown in Fig. 13. The energy-minimum-point operations are also feasible. For these beneficial features, the increase in the cell area could be acceptable depending on applications, e.g., mobile edge computing devices and neural network accelerators. In these applications, reducing leakage power is indispensable, and thus the ULVR-SRAMs have a great impact on these new applications [46]. A hybrid structure of ULVR and 6T cells could be also promising for caches in MPs and SoCs, which can reduce the area overhead, although a new architecture for the control needs to be developed.

VI. CONCLUSION

A design methodology of the ULVR-SRAM cell using HFPSs and its performance for PG are investigated. The ULVR-SRAM cell is comprised of fully-CMOS-based DM inverters, which can change its operational mode depending on V_{cell} controlled by the HFPSs. The modes are based on the NI and ST-inverter operations of the DM inverters. When the ordinary supply-voltage is applied to the cell, the cell can act as a high performance SRAM cell. When V_{cell} is reduced to an ultralow voltage, the cell can transition to the ULVR mode and dramatically reduce the leakage power without losing its data, i.e., the substantive PG can be achieved using the ULVR. The ability of leakage power reduction is enhanced by introducing the body biases that are automatically induced to the cell only during the ULVR mode by the help of an architecture using the HFPSs. The developed design methodology of the cell employs the QSNM for the ULVR mode as a design index. The transistor sizes and the bias condition for V_{cell} can be determined so as to minimize the leakage power with keeping a sufficiently high noise margin during the ULVR mode. An optimally designed ULVR-SRAM cell not only has excellent ability of leakage power reduction using the ULVR mode, but also show high performance on the normal SRAM operations that is comparable with conventional 6T cells. An 8KB ULVR-SRAM macro is also designed for analyzing the power and energy performances. The post-layout large-scale simulations of the optimally designed 8KB macro exhibits excellent PG ability: The leakage power can be reduced by $\sim$98% using the ULVR and a minimally short BET of 1.5μs can be achieved. The ULVR-SRAM can provide a new class of energy efficient PG architecture.

REFERENCES

[1] D. Harris and S. Harris, *Digital Design and Computer Architecture*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[2] M. Khellah *et al.*, "A 4.2GHz 0.3mm2 256kb dual-Vcc SRAM building block in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2006, pp. 2572–2581.

[3] S. Rusu *et al.*, "A 65-nm dual-core multithreaded Xeon processor with 16-MB L3 cache," *IEEE J. Solid-State-Circuits*, vol. 42, no. 1, pp. 17–25, Jan. 2007.

[4] N. Sakran, M. Yuffe, M. Mehalel, J. Doweck, E. Knoll, and A. Kovacs, "The implementation of the 65nm dual-core 64b merom processor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2007, pp. 106–107.

[5] J. B. Kuang *et al.*, "The design and implementation of a low-overhead supply-gated SRAM," in *Proc. 32nd Eur. Solid-State Circuits Conf. (ESSCIRC)*, Montreux, Switzerland, Sep. 2006, pp. 287–290.

[6] K. Ando *et al.*, "BRein memory: A single-chip binary/ternary reconfigurable in-memory deep neural network accelerator achieving 1.4 TOPS at 0.6 W," *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 983–994, Apr. 2018.

[7] A. Agrawal *et al.*, "Xcel-RAM: Accelerating binary neural networks in high-throughput SRAM compute arrays," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 8, pp. 3064–3076, Aug. 2019.

[8] X. Si *et al.*, "A dual-split 6T SRAM-based computing-in-memory unit-macro with fully parallel product-sum operation for binarized DNN edge processors," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 11, pp. 4172–4185, Nov. 2019.

[9] R. Puri, S. Leon, and B. Subhrajit, "Keeping hot chips cool," in *Proc. 42nd ACM Design Autom. Conf. (DAC)*, Anaheim, CA, USA, Jun. 2005, pp. 285–288.

[10] S. Rodriguez and B. Jacob, "Energy/power breakdown of pipelined nanometer caches (90nm/65nm/45nm/32nm)," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, Tegernsee, Germany, Oct. 2006, pp. 25–30.

[11] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.

[12] M. Anis and M. Elmasry, *Multi-Threshold CMOS Digital Circuits*, Norwell, MA, USA: Kluwer, 2003.

[13] Z. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban, H. Jacobson, and P. Bose, "Microarchitectural techniques for power gating of execution units," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, Newport Beach, CA, USA, Aug. 2004, pp. 32–37.

[14] Y. Kanno *et al.*, "Hierarchical power distribution with power tree in dozens of power domains for 90-nm low-power multi-CPU SoCs," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 74–83, Jan. 2007.

[15] V. George *et al.*, "Penryn: 45-nm next generation Intel core^{TM} 2 processor," in *Proc. IEEE Asian Solid-State Circuits Conf. (ASSCC)*, Jeju, South Korea, Nov. 2007, pp. 14–17.

[16] D. Apalkov, B. Dieny, and J. M. Slaughter, "Magnetoresistive random access memory," *Proc. IEEE*, vol. 104, no. 10, pp. 1796–1830, Aug. 2016.

[17] S. Bhatti, R. Sbiaa, A. Hirohata, H. Ohno, S. Fukami, and S. N. Piramanayagam, "Spintronics based random accessmemory: A review," *Mater. Today*, vol. 20, no. 9, pp. 530–548, Nov. 2017.

[18] P. F. Chiu *et al.*, "Low store energy, low VDDmin, 8T2R nonvolatile latch and SRAM with vertical-stacked resistive memory (memristor) devices for low power mobile applications," *IEEE J. Solid-State Circuits*, vol. 47, no. 6, pp. 1483–1496, Jun. 2012.

[19] D. Kitagata, S. Yamamoto, and S. Sugahara, "Design and energy-efficient architectures for nonvolatile static random access memory using magnetic tunnel junctions," *Jpn. J. Appl. Phys.*, vol. 58, Mar. 2019, Art. no. SBBB12.

[20] S. Mittal and J. S. Vetter, "A survey of software techniques for using non-volatile memories for storage and main memory systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 5, pp. 1537–1550, May 2016.

[21] J. Boukhobza, S. Rubini, R. Chen, and Z. Shao, "Emerging NVM: A survey on architectural integration and research challenges," *ACM Trans. Design Autom. Electron. Syst.*, vol. 23, no. 2, pp. 1–32, Nov. 2017.

[22] D. Kitagata, H. Yoshida, S. Yamamoto, and S. Sugahara, "Virtually nonvolatile retention SRAM cell using dual-mode inverters," in *Proc. IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, San Francisco, CA, USA, Oct. 2018, pp. 1–3.

[23] A. Agarwal, H. Li, and K. Roy, "A single-Vt low-leakage gated-ground cache for deep submicron," *IEEE J. Solid-State Circuits*, vol. 38, no. 2, pp. 319–327, Feb. 2003.

[24] A. J. Bhavnagmala, S. V. Kosonocky, M. Immediato, D. Knebel, and A.-M. Haen, "A pico-joule class, 1 GHz, 32 KJ3yte x 64b DSP SRAM with self reverse bias," in *Proc. Symp. VLSI Circuits*, Kyoto, Japan, Jun. 2003, pp. 251–252.

[25] P. Elakkumanan, A. Narasimhan, and R. Sridhar, "NC-SRAM -a low-leakage memory circuit for ultra deep submicron designs," in *Proc. IEEE Int. SOC Conf.*, Portland, OR, USA, Sep. 2003, pp. 3–6.

[26] M. Khellah *et al.*, "A 256-Kb dual-V_{CC} SRAM building block in 65-nm CMOS process with actively clamped sleep transistor," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 233–242, Jan. 2007.

[27] H. Makino *et al.*, "An auto-backgate-controlled MT-CMOS circuit," in *Proc. Symp. VLSI Circuits*, Honolulu, HI, USA, Jun. 1998, pp. 42–43.

[28] K. Nii *et al.*, "A low power SRAM using auto-backgate-controlled MT-CMOS," in *Proc. Int. Symp. Low Power Electron. Design (ISLPED)*, Monterey, CA, USA, Aug. 1998, pp. 293–298.

[29] L. Chang *et al.*, "Stable SRAM cell design for the 32 nm node and beyond," in *Proc. Symp. VLSI Circuits*, Kyoto, Japan, Jun. 2005, pp. 128–129.

[30] Y. Morita *et al.*, "An area-conscious low-voltage-oriented 8T-SRAM design under DVS environment," in *Proc. Symp. VLSI Circuits*, Kyoto, Japan, Jun. 2007, pp. 256–257.

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

[32] H. Noguchi *et al.*, "A 10T non-precharge two-port SRAM for 74% power reduction in video processing," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, Porto Alegre, Brazil, Mar. 2007, pp. 107–112.

[33] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

[34] S. Jain *et al.*, "A 280mV-to-1.2V wide-operating-range IA-32 processor in 32nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2012, pp. 66–68.

[35] G. Pasandi and S. M. Fakhraie, "An 8T low-voltage and low-leakage half-selection disturb-free SRAM using bulk-CMOS and FinFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 7, pp. 2357–2363, Jul. 2014.

[36] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV robust Schmitt trigger based subthreshold SRAM," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2303–2313, Oct. 2007.

[37] J. P. Kulkarni and K. Roy, "Ultralow-voltage process-variation-tolerant schmitt-trigger-based SRAM design," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 2, pp. 319–332, Feb. 2012.

[38] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "A variation-tolerant sub-200 MV 6-T subthreshold SRAM," *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2338–2348, Oct. 2008.

[39] *VLSI Design and Education Center: VDEC, The University of Tokyo, Japan*. Accessed: Feb. 15, 2021. [Online]. Available: http://www.vdec.u-tokyo.ac.jp/English/index.html

[40] H. Kawaguchi, K. Nose, and T. Sakurai, "A CMOS scheme for 0.5 V supply voltage with pico-ampere standby current," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 1998, pp. 192–193.

[41] S. Ohbayashi *et al.*, "A 65-nm SoC embedded 6T-SRAM designed for manufacturability with read and write operation stabilizing circuits," *IEEE J. Solid-State Circuits*, vol. 42, no. 4, pp. 820–829, Mar. 2007.

[42] H. Makiyama *et al.*, "Suppression of die-to-die delay variability of silicon on thin buried oxide (SOTB) CMOS circuits by balanced P/N drivability control with back-bias for ultralow-voltage (0.4 V) operation," in *IEEE Int. Electron Devices Meeting (IEDM) Dig. Tech Papers*, Washington, DC, USA, Dec. 2013, pp. 822–825.

[43] R. Jain *et al.*, "A 0.45-1V fully integrated reconfigurable switched capacitor step-down DC–DC converter with high density MIM capacitor in 22nm tri-gate CMOS," in *Proc. Symp. VLSI Circuits*, Kyoto, Japan, Aug. 2013, pp. 174–175.

[44] N. Sugii, R. Tsuchiya, T. Ishigaki, Y. Morita, H. Yoshimoto, and S. Kimura, "Local V_{th} variability and scalability in silicon-on-thin-box (SOTB) CMOS with small random-dopant fluctuation," *IEEE Trans. Electron Devices*, vol. 57, no. 4, pp. 835–845, Apr. 2010.

[45] E. Karl *et al.*, "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active V_{MIN}-enhancing assist circuitry," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2012, pp. 230–238.

[46] Y. Shiotsu, S. Yamamoto, and S. Sugahara, "Proposal and performance prediction of a BNN accelerator using ULVR-SRAM," in *Proc. 82nd JSAP Autumn Meeting*, Aichi, Japan, Sep. 2021, paper no. 12p-N304-8.