A 10 nm FinFET 128 Mb SRAM With Assist Adjustment System for Power, Performance, and Area Optimization

Taejoong Song, Woojin Rim, Sunghyun Park, Yongho Kim, Giyong Yang, Hoonki Kim, Sanghoon Baek, Jonghoon Jung, Bongjae Kwon, Sungwee Cho, Hyuntaek Jung, Yongjae Choo, and Jaeseung Choi

***Abstract*— Two 128 Mb 6T SRAM test chips are implemented in a 10 nm FinFET technology. A 0.040 μm^2 6T SRAM bitcell is designed for high density (HD), and 0.049 μm^2 for high performance (HP). The various SRAM assist schemes are explored to evaluate the power, performance, and area (PPA) gain, and the figure-of-merit (FOM) is induced by the minimum operating voltage (V_{MIN}) and assist overheads. The dual-transient wordline scheme is proposed to improve the V_{MIN} by 47.5 mV for the 128 Mb 6T-HP SRAM. The suppressed bitline scheme with negative bitline improves the V_{MIN} by 135 mV for the 128 Mb 6T-HD SRAM. The FOM of PPA gain evaluates the optimum SRAM assist for the different bitcells based on the applications.**

***Index Terms*— 10 nm FinFET, dual-transient wordline (DTWL), figure of merit (FOM), high-density (HD), high-performance (HP), low-power, power, performance, and area (PPA) gain, SRAM assist.**

I. Introduction

THE operating voltage (V_{OP}) of logic transistors has been reduced over technologies to provide low power. Then, to meet the high performance (HP), the threshold voltage (V_{TH}) has also been reduced accordingly. Fig. 1 illustrates the trend of V_{OP} and V_{TH} over technologies [1]–[16]. However, V_{TH} reduction does not meet the trend of V_{OP} reduction due to the process limitation of gate engineering. Furthermore, since V_{TH} is a dependent variable of the statistical distribution at a nanoscale process, it causes a diminution of the voltage headroom ($V_{OP} - V_{TH}$) as shown in Fig. 1. Thus, the logic gate immunity is decreased against the noise and variation by the gradual decrease of V_{TH} [6]–[8]. In order to improve the stability in a low-voltage region, the innovative techniques, such as a postprocessing architecture for error correction or a statistical timing analysis, have been used in [17]. However, it is not easy for an SRAM bitcell to recover from the small voltage headroom, since the transistors of an SRAM bitcell paradoxically support the stability and writability [18].

Manuscript received April 25, 2016; revised June 21, 2016 and August 27, 2016; accepted September 1, 2016. Date of publication October 28, 2016; date of current version January 4, 2017. This paper was approved by Guest Editor Atsushi Kawasumi.

The authors are with Samsung Electronics Company, Ltd., Hwaseong 445-701, South Korea, (e-mail: tj.song@samsung.com; woojin.rim@samsung.com).

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

Digital Object Identifier 10.1109/JSSC.2016.2609386

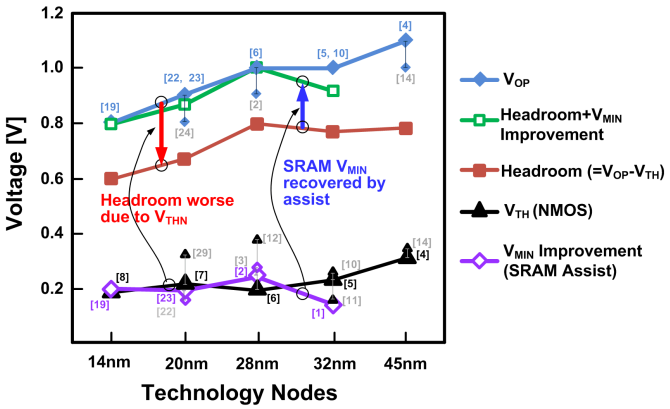


Fig. 1. Trend of the operating voltage (V_{OP}), threshold voltage (V_{TH}), and the voltage headroom (V_{OP}–V_{TH}) over technologies [1]–[16]. The voltage headroom is reduced by the saturated V_{TH}, while the SRAM voltage headroom is compensated by V_{MIN} improvement.

Recent introduction of assist schemes help to recover the challenges of SRAM voltage-headroom. Fig. 1 describes the minimum operating voltage (V_{MIN}) of an SRAM that is improved by an assist. The improved headroom is dotted as "$V_{OP} - V_{TH} + V_{MIN}$ improvement," which is similar to the trend of the V_{OP} trend. It is meaningful that assist helps SRAM voltage-headroom to trace the trend of V_{OP}. Therefore, system-on-chip (SoC) designers can lower the V_{OP} of both SRAM and logic for low-power applications in the nanoscale technology. Meanwhile, an SRAM assist requires the timing or area overhead for the power gain, thus faltering the total gain of power, performance, and area (PPA). This paper explores the various SRAM assist techniques that provide the optimum PPA gain according to the applications in a 10 nm FinFET technology. Then, the dual-transient wordline (DTWL) is proposed to achieve the best PPA gain for a specific bitcell type. The rest of this paper is organized as follows. Section II introduces the 10 nm FinFET technology and the 6T SRAM bitcell. Section III explains the conventional SRAM assist techniques. Section IV illustrates the DTWL technique with the challenge of metal resistance for an assist. Section V shows the SRAM macro with the various assist techniques. Section VI describes the figure-of-merit (FOM) of PPA gain with an SRAM assist. Section VII explains the implementation and measurement of a test chip.

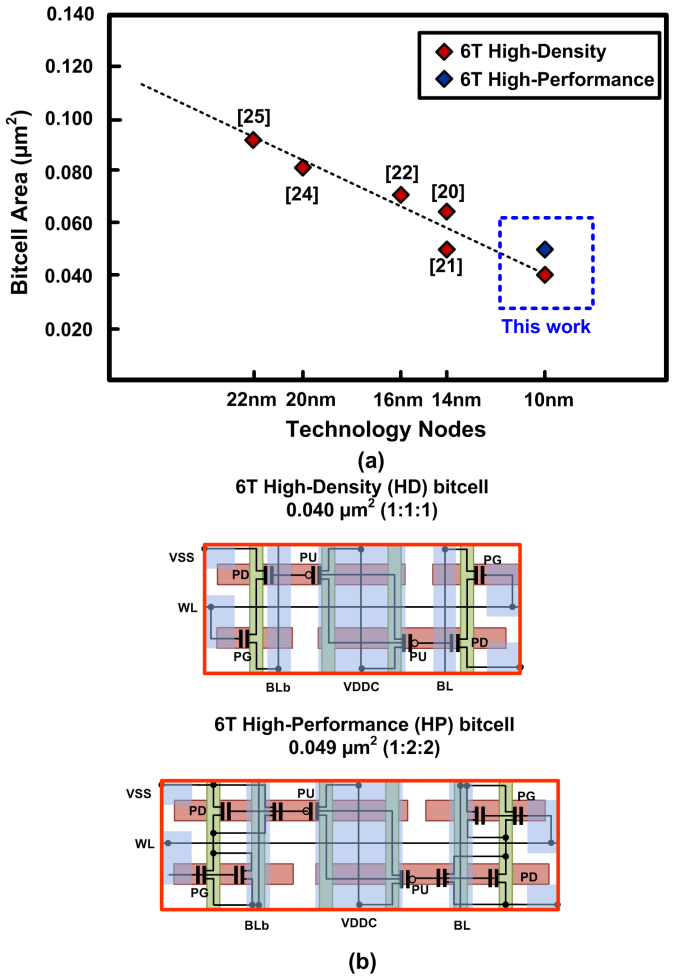


Fig. 2. (a) Scaling trend of 6T SRAM bitcell over technologies. (b) 10 nm FinFET 6T SRAM bitcells of HD and HP.

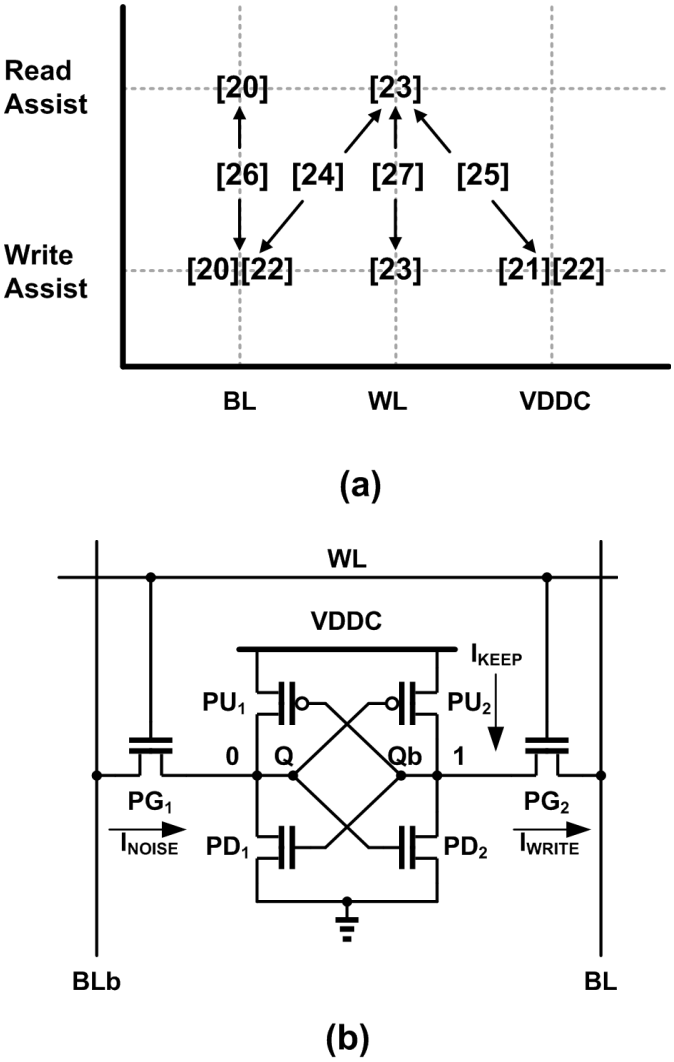


Fig. 3. (a) Conventional SRAM assist techniques, which are categorized into the assist knobs of WL, BL, and cell power (VDDC). (b) 6T SRAM bitcell schematic.

II. 10 nm FinFET Technology

The developed 10 nm FinFET logic technology features a 64 nm critical poly pitch and a 48 nm metal pitch [19] that enables gate engineering for HP and low power. In addition, the smallest SRAM bitcell is developed with the aid of the innovative 10 nm FinFET technology. Fig. 2(a) illustrates the scaling trend of a 6T SRAM bitcell over technologies. Compared with the 14 nm FinFET SRAM bitcell [20], the 10 nm FinFET SRAM bitcell demonstrates an area reduction by 38%, the highest density published thus far [20]–[25]. Fig. 2(b) illustrates the 0.040 μm^2 6T high density (HD) SRAM bitcell with 1:1:1 fins for pull-up (PU), pass-gate (PG), and pull-down. The 0.049 μm^2 6T HP SRAM bitcell is designed with 1:2:2 fins. The 6T-HP bitcell leverages the high-speed feature of 1:2:2 fins with a 22% larger area versus the 6T-HD bitcell. Using a mix of HD and HP bitcells, SoC designer optimizes the PPA based on the application. Then, the additional PPA gain using a small bitcell is achieved by manipulating appropriate SRAM assist techniques.

III. SRAM Assist Schemes

The conventional SRAM assist schemes are categorized based on assist knobs that include WL, bitline (BL), and cell-power (VDDC). Fig. 3 illustrates the conventional SRAM assist techniques for write-assist (WA) and read-assist (RA) with a 6T SRAM bitcell. Intrinsically, the read stability of an SRAM bitcell is distorted by the noise from the PG transistor, and therefore, the read margin is improved by controlling BL or WL. BL is discharged [20] or suppressed BL (SBL) [26] to reduce the half-selected BL noise (I_{NOISE}). Similarly, WL is lowered (WLUD) to reduce the I_{NOISE} [23]–[25], [27]. These schemes allow a slow and safe BL evaluation with the bitcell data. However, WLUD requires additional timing for the safe operation, thus decreasing the write margin with a smaller I_{WRITE}. Meanwhile, because a write margin is hurt by the strong PU current (I_{KEEP}) and a weak PG current (I_{WRITE}), VDDC, WL, and BL are efficiently controlled to ensure that I_{KEEP} remains small and I_{WRITE} large. VDDC is collapsed (VDDCUD) to lower the PU PMOS strength [21], [22], [25]. The VDDCUD does not impact the half-selected bitcells as that of the active bitcell; however, it causes bitcell instability of the unaccessed bitcells in the same column. BL is negatively-boosted [negative BL (NBL)]

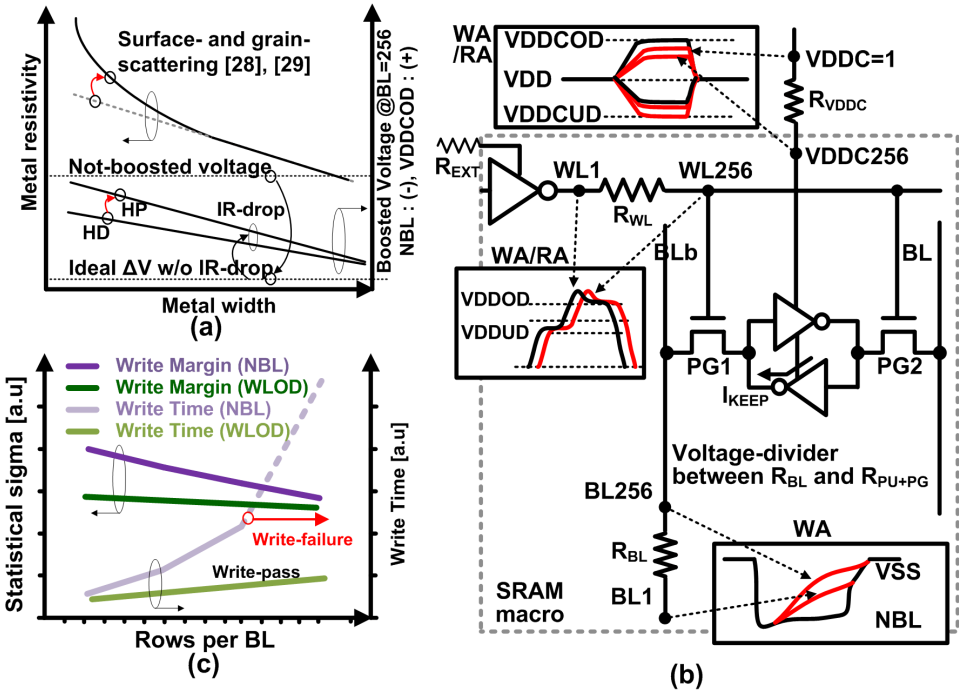


Fig. 4. (a) Resistance trend versus metal width. As metal pitch and width decrease, metal resistance increases more rapidly due to surface and grain scattering. (b) Conceptual diagram of the resistance impact on the SRAM assist schemes. (c) Resistance impact on the write margin and the timing for WLOD and NBL.

to improve the writability of the selected bitcell [20], [22], [24], [26], thus paving way to create the timing and area overhead for the NBL. Further, as the V_{OP} of an SRAM macro reduces, the more difficult it is to meet the V_{MIN} using single assist of RA or WA. Therefore, the mixed SRAM assists are adopted [24]–[27] that utilize the features of both RA and WA at the same time. NBL is applied to recover the weak writability of WLUD [24]. Similarly, VDDCUD is added to improve the write margin which is damaged by WLUD [25]. SBL with NBL manages a BL voltage to maintain an optimum write and read margin [26]. Meanwhile, WLUD with WLOD is applied with the timing interval [27] to improve the read margin and recovering write margin consecutively. In this paper, we present the mixed assist of the WLUD with WLOD scheme that is similar to [27] for the assist knob. However, the proposed one shows the differentiation using the external supply voltages to minimize the timing and area penalty.

IV. Dual-Transient Wordline Scheme

In a 10 nm FinFET technology, metal resistance is the key factor in choosing an SRAM assist. Fig. 4(a) illustrates the resistance trend versus the metal width. Decrease in the metal pitch and width leads to a proportional increase in the resistance. Nevertheless, resistance increases more rapidly than the metal width reduction due to surface and grain scattering [28], [29]. Increase in the resistance values beyond a certain level prevents the assist from helping the write operation. Fig. 4(b) illustrates the resistance impact on an SRAM assist. Since VDDC or NBL drives the source of PMOS or NMOS, respectively, the boosted voltage level is contaminated by a large resistance, which is impacted by the voltage divider between metal resistance and bitcell transistors. For example,

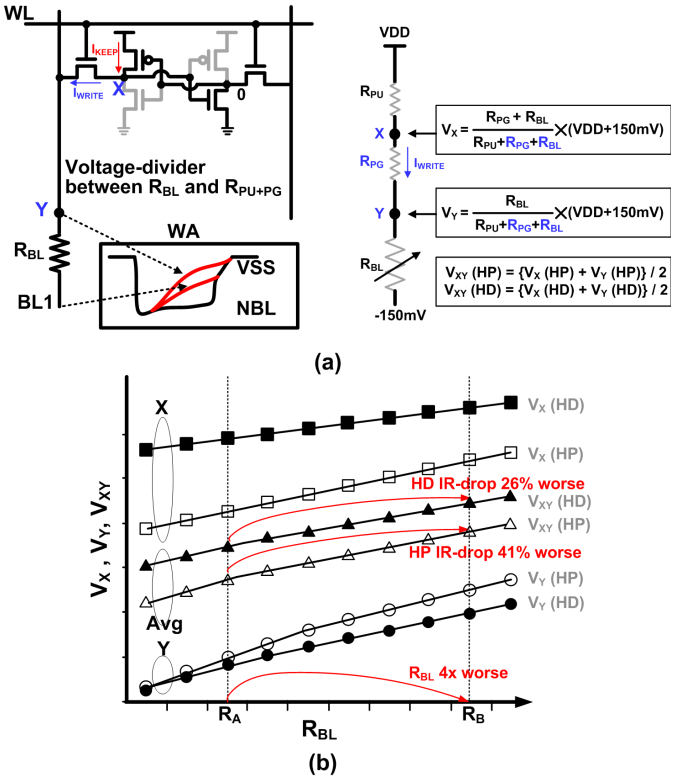


Fig. 5. BL voltage is impacted by IR drop due to the PU resistance (R_{PU}), PG resistance (R_{PG}), and BL resistance (R_{BL}). (a) IR drop is decided by voltage divider between R_{PU}, R_{PG}, and R_{BL}. (b) As R_{BL} increases, V_{XY} (HP) is impacted more than V_{XY} (HD) with IR drop.

the negative voltage with NBL is driven from BL1 to BL256. Since BL256 node is kept by I_{KEEP} through PU and PG transistors, the voltage level of BL256 is set by the voltage divider

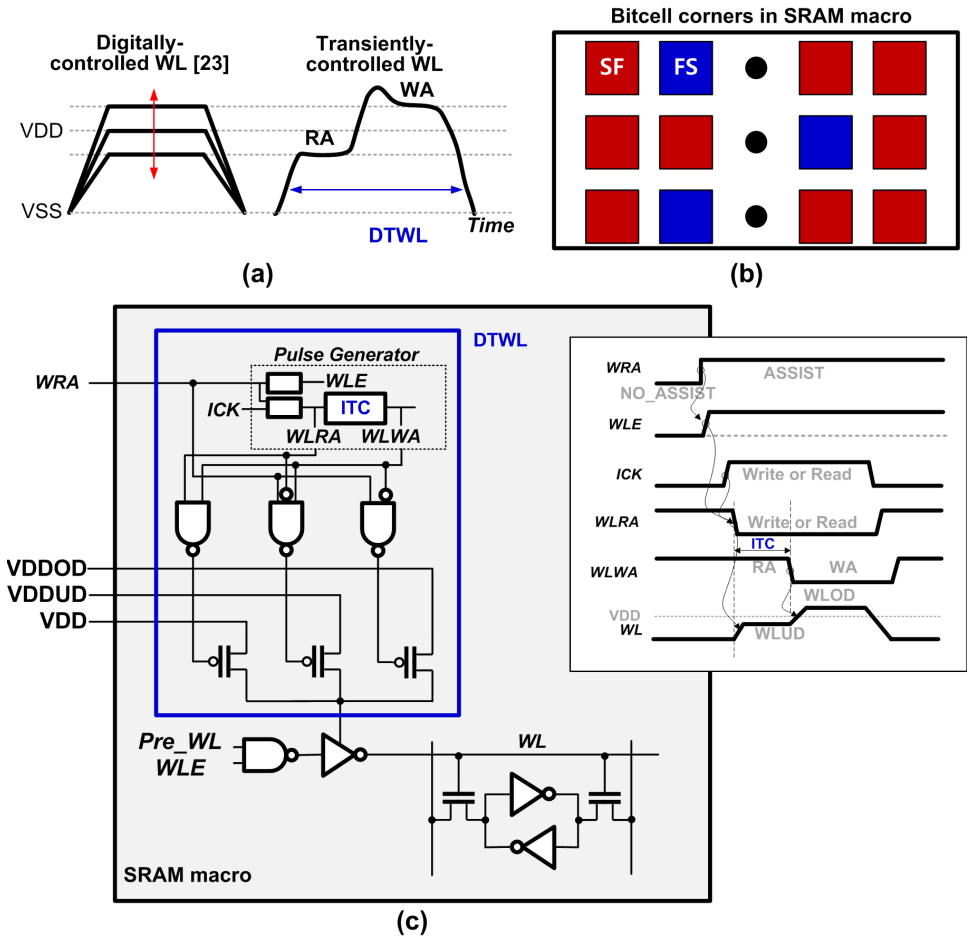


Fig. 6. (a) Timing diagram of the digitally controlled WL [23] and the proposed DTWL. (b) Bitcell process corners in an SRAM macro. (c) DTWL schematic with the ITC circuit.

between BL resistance (R_{BL}), the resistance of PU (R_{PU}), and PG (R_{PG}). If R_{BL} is too large not to overcome $R_{\mathrm{PU+PG}}$, BL256 cannot be driven below the write voltage of V_{TRIP}, where a bitcell is flipped to a different data. Fig. 4(c) illustrates the impact of resistance on a write margin and timing with an NBL in a 10 nm FinFET technology. When rows-per-BL (RPB) increases over a certain level in an SRAM macro, the NBL cannot aid the write operation and results in a write failure. Moreover, the write failure due to a large resistance cannot be improved with additional timing delay. Meanwhile, since WLOD drives the gate instead of the source of the PG transistor, it handles large resistance more easily. Write margin is recovered with timing delay in the case of large metal resistance. With a small resistance of BL, NBL is more effective than WLOD to improve a write margin, because NBL boosts the drain-source (V_{DS}) and the gate-source (V_{GS}) of PG, while WLOD boosts the V_{GS} of PG. However, NBL stops to improve a write margin over a certain RPB due to large resistance, while WLOD is still effective to improve the write margin. Write margin is compared between WLOD and NBL over RPB in Fig. 4(c). For a small RPB, NBL is more effective to improve the write margin, but NBL stops to improve the write margin for large RPB. Therefore, WLOD needs to be chosen for large RPB. Otherwise, an SRAM macro should be designed with a smaller RPB architecture, which makes the area and timing overhead. Fig. 5 shows the IR drop impact with a large BL resistance for 6T-HP and HD bitcells. The voltage level of X and Y are calculated by considering the IR drop due to R_{PU}, R_{PG}, and R_{BL}. Then, V_{XY} is described by the average voltage of V_X and V_Y to consider the voltage level at the rising time of WL as shown in Fig. 5(a). Fig. 5(b) describes the IR drop impact of V_X, V_Y, and V_{XY} over R_{BL}. V_{XY} (HP) is lower than V_{XY} (HD) intrinsically, since R_{PG} is smaller to connect X to BL more closely. However, the impact of IR drop with large R_{BL} gets worse for 6T-HP than HD. For example, when R_{BL} increases from R_A to R_B by 4×, V_{XY} (HD) gets worse by 26%. Meanwhile, V_{XY} (HP) shows a 41% impact, which is 11% worse than V_{XY} (HD). Therefore, even if 6T-HP shows the better writability than 6T-HD with the smaller R_{PG}, it suffers from IR drop more easily. To avoid the resistance impact of BL, a WLOD or a WLUD is selectively applied at the different process corners for WA and RA [23]. As illustrated in Fig. 6(a), a WLOD is applied to improve the write margin in an NMOS Slow and PMOS Fast (SF) corner, and the WLUD is applied to an NMOS Fast and PMOS Slow (FS) to improve the read margin. However, since all the bitcells in an SRAM macro are not processed under the same corner as shown in Fig. 6(b), the single WL voltage does

not improve the margin of the bitcells in different process corners. The WLUD degrades the write margin of the bitcell in an SF corner, while the WLOD hurts the read margin of the bitcell in an FS corner. In this paper, the DTWL is proposed to adopt the features of both WLUD and WLOD. Fig. 6(c) illustrates the schematic and timing diagram of DTWL. The DTWL leverages the external supply voltages to minimize the timing and area overheads. The DTWL leverages the external supply voltage to minimize the timing and area overheads. Since the external supply voltages are generated from a low-dropout (LDO) regulator, its overhead needs to be considered for PPA calculation with SRAM assist. However, in the conventional SoC system with thousands of SRAM modules, the overheads of the LDO are minimized by sharing the external LDOs for thousands of SRAM modules. Therefore, the area overheads due to the external voltages sources are negligible in the SoC applications. Then, the LDO for SRAM assist is considered to provide the tolerant external voltage compared with the normal supply voltage. Therefore, the overhead of the external supply voltage for SRAM assist is not seriously considered in this literature. Then, by applying WA and RA sequentially, it overcomes the risk of the predefined single assist scheme. The DTWL is driven by three supply voltages, namely, VDD, VDDUD (= VDD − ΔV), and VDDOD (= VDD + ΔV). In a normal operation without WA or RA, VDD is used as the WL supply voltage. Compared with the digitally controlled WL scheme [23], VDDUD is driven for RA followed by the VDDOD for WA with timing delay. Then, the optimum delay is controlled by the internal-timing-control (ITC). The ITC is tuned recursively after all the bitcells are tested for RA and WA in an SRAM macro. The ITC timing span ensures the stability of the weak bitcell that is different according to the process corners.

The DTWL has the drawback of timing overhead like WLUD. However, the penalty is partially recovered by the consecutive assist of WLOD. Fig. 7(a) illustrates the timing diagram of write operation at NN and SF corners. In an NN corner, the bitcell data of Q and Qb are flipped before WLOD is enabled. Contrarily, in an SF corner, bitcell write margin is too small to create a safe write operation during WLUD. Therefore, WLOD is necessary to flip the bitcell data. Fig. 7(b) illustrates the timing overhead of WLUD and DTWL. When DTWL applies WLOD after WLUD, it recovers a 47% timing overhead by WLOD, which is degraded by 62% due to WLUD at a specific ITC delay. Since the write timing is decided at the worst corner of SF, the timing penalty is considered in an SF corner. Therefore, DTWL helps to reduce the overall timing overhead considering the worst process corner. The DTWL is applied at an SF process corner to help write margin, which recovers the timing loss with WLUD. At an NN process corner, the WLOD is not necessary, since WA is not necessary. Therefore, an SRAM macro is thoroughly tested for all the bitcells to check the weak bitcell for RA, WA, or the mixed assist.

V. SRAM Macro Design With Assist

The various SRAM assist schemes are explored in a 10 nm FinFET technology. Fig. 8 illustrates the SRAM

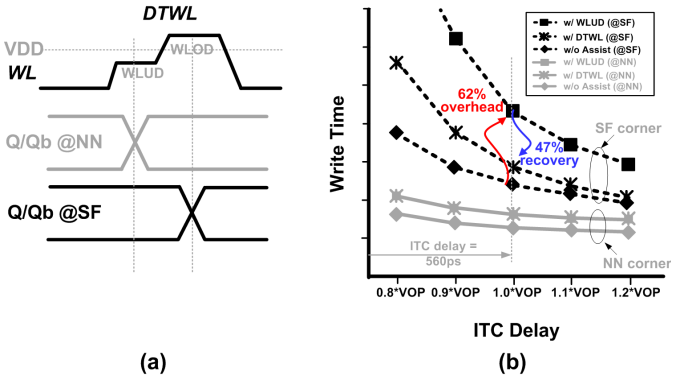


Fig. 7. (a) Timing diagram of write operation at NN and SF corner. Write operation happens before WLOD at NN process, and in WLOD at SF process. (b) Timing overhead of WLUD and DTWL. The DTWL recovers the timing overhead of WLUD by WLOD.

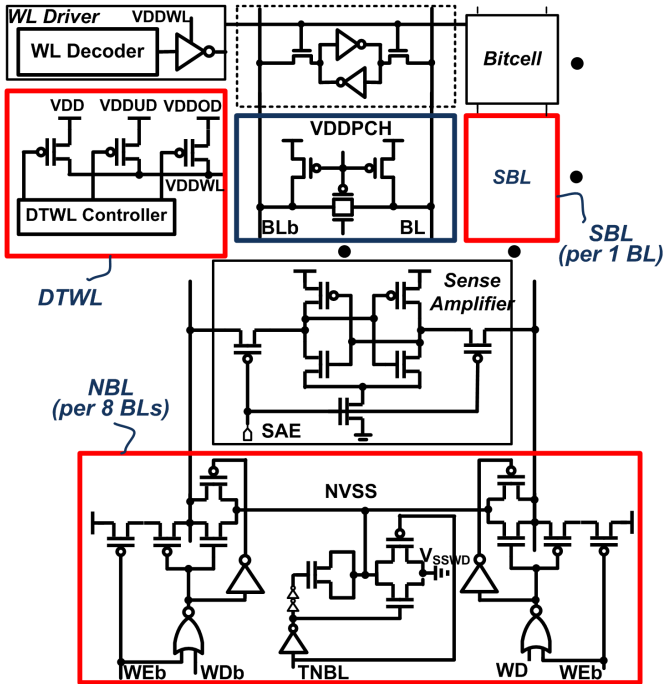


Fig. 8. SRAM macro with the various assist schemes. Assist types for WA, RA, and the mixed assist with WA and RA.

macro with the SRAM assist schemes. The proposed DTWL is implemented in a WL driver, which selects WLUD or WLOD through the ITC as explained in Section IV. SBL is implemented in a precharge block to create a lower precharge voltage level. NBL is implemented in a write buffer, which generates the negative voltage through an internal charge pump. Table I describes the SRAM assist features that are tested in the SRAM macro. NBL or WLOD is tested for WA, while SBL or WLUD is explored for RA. Then, three types of mixed assist schemes are tested to maximize the assist gain, which include the assist schemes of SBL with NBL, DTWL, and WLUD with NBL. Fig. 9 illustrates 528 kb (= 512 kb + 16 kb) SRAM subarray with 16 33 kb unit arrays, which are configured with 264 columns 128 rows. Fig. 10 illustrates the 128 Mb 6T SRAM test chip,

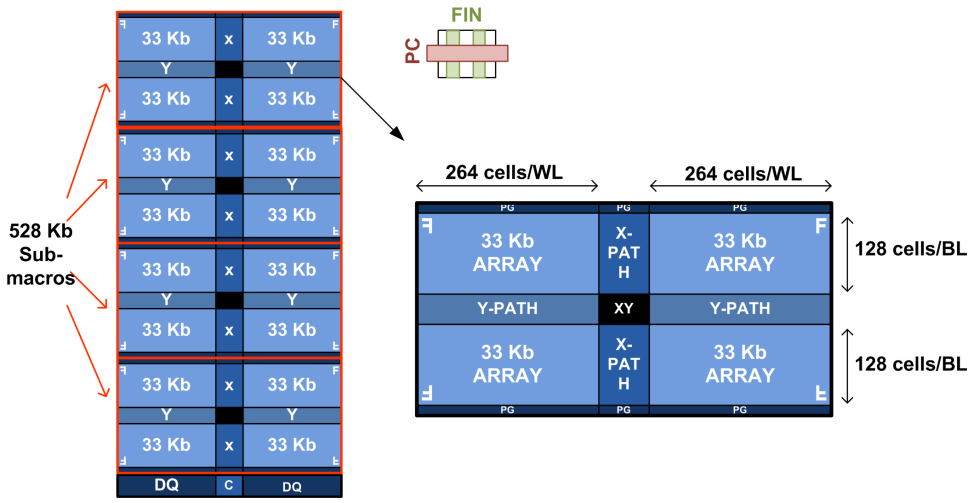


Fig. 9. 528 kb (= 512 kb + 16 kb) SRAM subarray with 16 × 33 kb unit arrays.

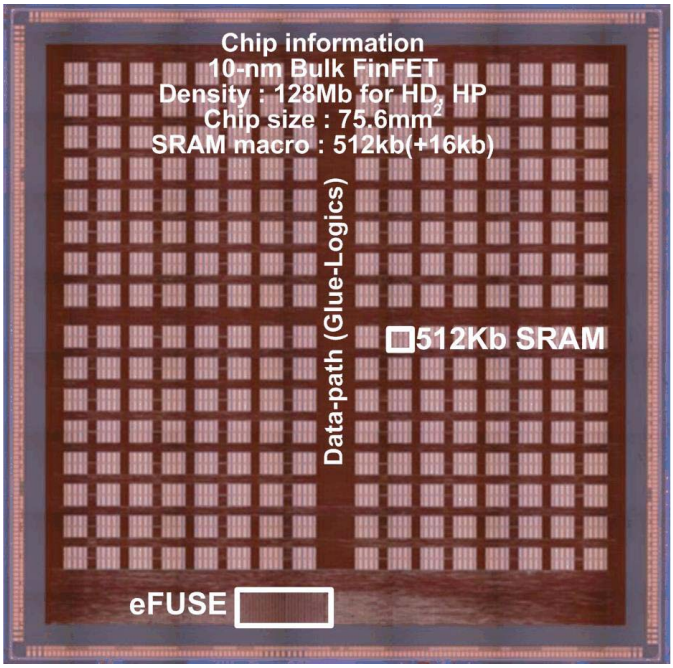


Fig. 10. 10 nm FinFET 128 Mb 6T SRAM test chip. The 6T-HD and HP are designed with the same footprint for the test chip.

which is designed by 256 × 512 kb SRAM macros. A 10 nm FinFET eFUSE macro is used to repair the failed bitcells. The 6T-HD and HP have the same footprint with similar chip architectures.

VI. Figure of Merit of PPA Gain

As described in Section III, an SRAM assist improves V_{MIN} for low power. However, it is not easy to find the optimum SRAM assist, since the power gain is diminished by the speed and area penalty in perspective of PPA gain. Therefore, in this section, the FOM of the PPA gain is induced for the optimum SRAM assist. First, the weighting factors of PPA are extracted from the features of the bitcell. When the 6T-HP bitcell is

TABLE I
SRAM Assist Types

AST	Write-ability	Read-stability	Assist Scheme
1	↑		*NBL*
2	↑		*WLOD*
3		↑	*SBL*
4		↑	*WLUD*
5	↑	↑	*SBL+NBL*
6	↑	↑	*DTWL*
7	↑	↑	*WLUD+NBL*

selected in the SoC design, it is intended to utilize the high speed by sacrificing the area and power gain, comparing the 6T-HD bitcell and vice versa. Fig. 11 illustrates the relative PPA of the 6T-HD and HP bitcell. The area factor (f_a) and power factor (f_p) are the relative area and power gain of 6T-HD bitcell versus 6T-HP bitcell, respectively. The speed factor (f_s) is the speed gain of 6T-HP bitcell versus 6T-HD bitcell. These factors are converted into the weighting factors of W_P, W_S, and W_A for 6T-HD as follows:

$$\begin{aligned} \text{SUM (HD)} &= (1/f_P) + 1 + (1/f_a) \\ W_P\ \text{(HD)} &= (1/f_P)/\text{SUM} \\ W_S\ \text{(HD)} &= 1/\text{SUM} \\ W_A\ \text{(HD)} &= (1/f_a)/\text{SUM}. \end{aligned}$$

Similarly, the weighting factors for 6T-HP are defined as shown in Table II.

Second, the power, speed, and area gain are extracted from the features of an SRAM macro with an assist. Then, by applying the weighting factors to PPA, the PPA_{eff} is

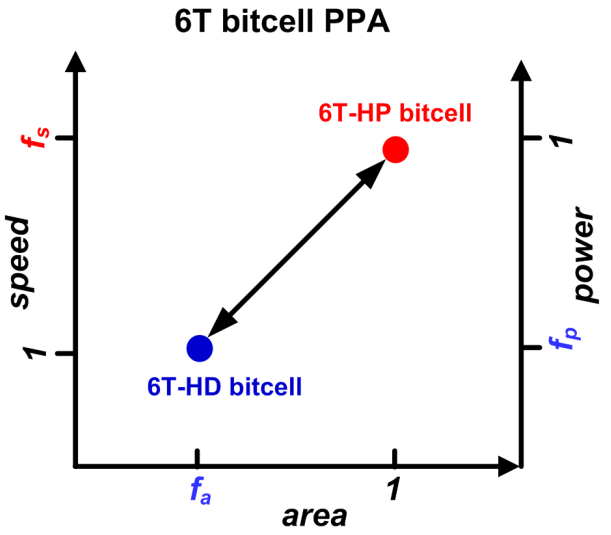


Fig. 11. Relative PPA of 6T-HD and HP SRAM bitcell. The weighting factors are extracted from the relative PPA of bitcells. The FOM is calculated with the PPA of SRAM macro and the weighting factors of bitcell PPA.

TABLE II
FOM OF PPA GAIN

		HD	HP
bitcell PPA	*power*	f_p	*1*
	speed	*1*	f_s
	area	f_a	*1*
	SUM	$1/f_p + 1 + 1/f_a$	$1 + f_s + 1$
Weighting Factor	W_P	$(1/f_p)$ / SUM	1 / SUM
	W_S	1 / SUM	f_s / SUM
	W_A	$(1/f_a)$ / SUM	1 / SUM
PPA_{eff}		(1-Power)×W_P + (Speed-1)×W_S + (1-Area)×W_A + 1	

induced as

$$\text{PPA}_{\text{eff}} = (1 - \text{Power}) \times W_P + (\text{Speed} - 1) \times W_S + (1 - \text{Area}) \times W_A + 1$$

where *Power*, *Speed*, and Area represent the PPA gain of an SRAM macro with an assist; W_P denotes the weighting factor of the bitcell power; W_S denotes the weighting factor of the bitcell speed; and W_A denotes the weighting factor of the bitcell area, which is summarized in Table II.

For example, if WLUD with NBL is used in 6T-HP, V_{MIN} is improved by 220 mV. It reduces the power by 32%, which considers the dynamic power and leakage power. Then, the WLUD with NBL assist requires additional 26% timing overhead due to the small I_{WRITE} and 6% area overhead with the internal charge pump. Finally, the PPA_{eff} of the 6T-HP is calculated as

$$\text{PPA}_{\text{eff}} = (1 - 0.68) \times 0.25 + (0.74 - 1) \times 0.50 + (1 - 1.06) \times 0.25 + 1 = 0.93.$$

This implies that WLUD with NBL helps V_{MIN} by enabling a 32% low power. However, when the timing and area overheads are considered, the PPA_{eff} is less than unity. In this way, the PPA_{eff} is calculated to understand the PPA gain of an SRAM assist with different SRAM assist schemes.

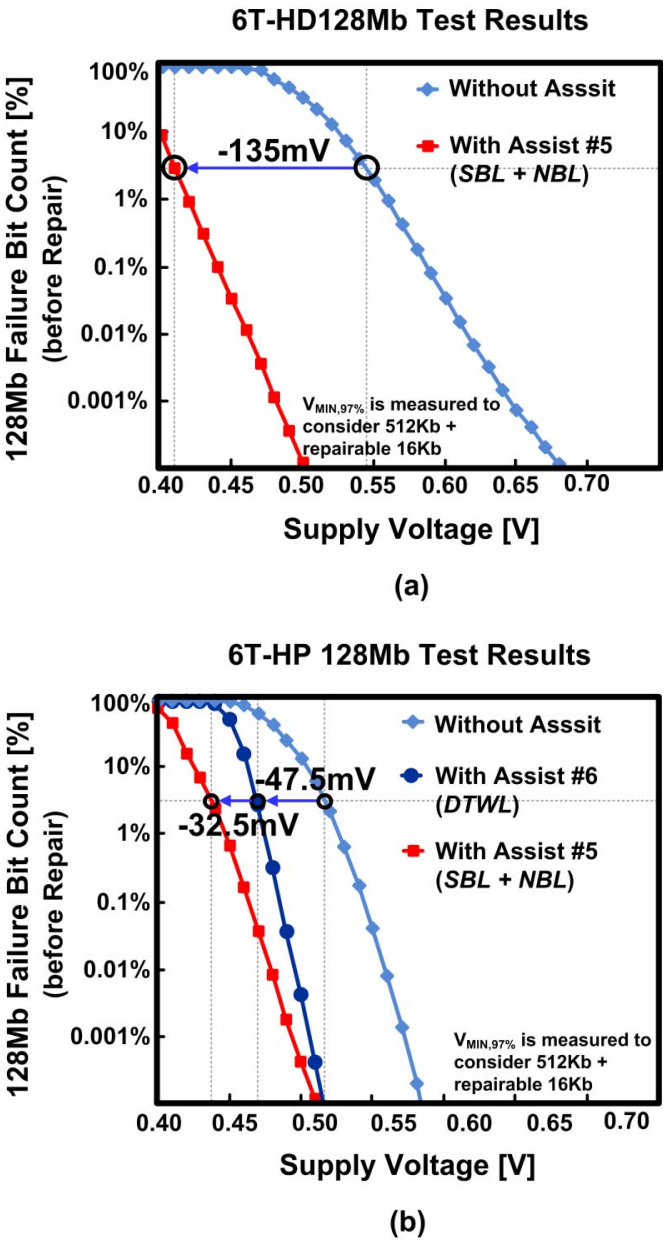


Fig. 12. V_{MIN} measurement result of 128 Mb SRAM test chips with WA and RA of (a) 6T-HD and (b) 6T-HP bitcell. The V_{MIN} is measured at a 97% yield as V_{MIN}, 97% by considering the redundancy.

VII. EXPERIMENTAL RESULTS

The 10 nm FinFET 6T SRAM test chip is tested to explore the SRAM assist. Fig. 12 demonstrates the V_{MIN} measurement results for 6T-HD and 6T-HP. V_{MIN} is measured at a 97% yield, which is described as V_{MIN}, 97%. The 97% yield is selected to consider 512 kb bitcells with 16 kb repairable bitcells in the 128 Mb SRAM. Therefore, the failure bitcells are assumed to be repaired ideally up to 3%. The SBL with NBL scheme improves the 6T-HD V_{MIN}, 97% by 135 mV, and the 6T-HP by 80 mV in comparison without the assist scheme.

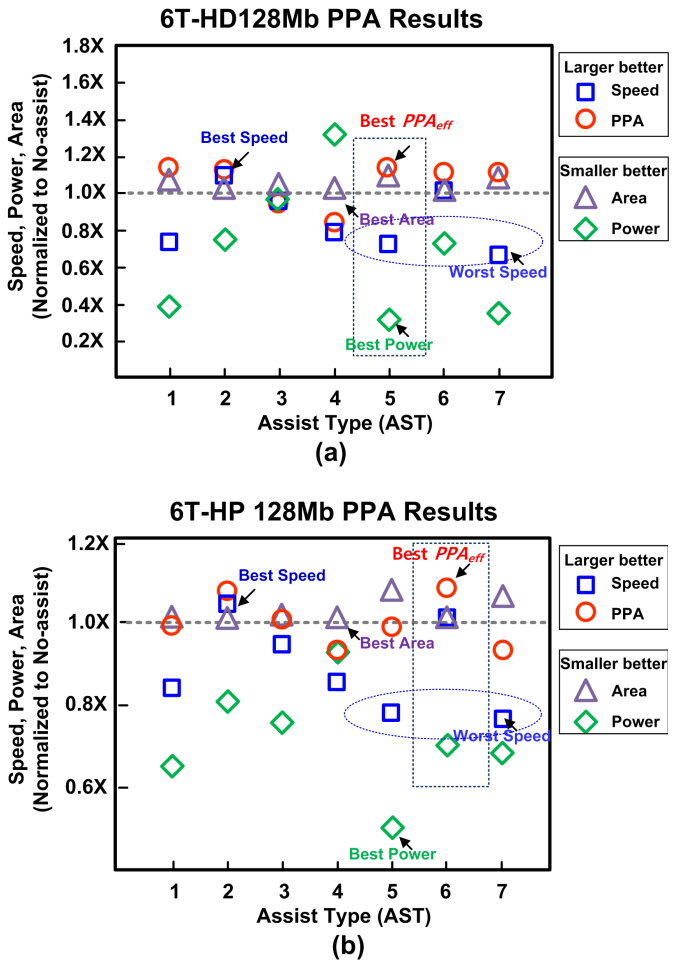


Fig. 13. PPA results of 128 Mb SRAM test chips with WA and RA of (a) 6T-HD and (b) 6T-HP bitcell. The PPA_{eff} is calculated by power gain with V_{MIN} improvement and the speed and area overhead of SRAM assist schemes.

Then, the proposed DTWL improves the 47.5 mV V_{MIN}, 97% from without assist. Based on the experiment, an SRAM macro is designed to use the low V_{MIN} for low power. However, it costs the speed and area overhead due to the SRAM assists. Therefore, the PPA_{eff} must be extracted to understand the effective PPA gain of an SRAM assist as described in Section VI.

Fig. 13(a) illustrates the PPA with the various SRAM assist schemes for 6T-HD. The assist types are described in Table I. The PPA_{eff} is calculated with the power reduction and the overheads of an SRAM assist. The 6T-HD illustrates the best speed gain for WLOD (AST 2), since WLOD provides a higher I_{WRITE} with less timing penalty. WLUD (AST 4) creates the smallest area overhead by utilizing the external supply voltage directly. WLUD with NBL (AST 7) demonstrates the slowest speed, since it requires additional timing to create the negative voltage for WA, and WLUD degrades the performance additionally. Finally, the SBL with NBL (AST 5) demonstrates the best PPA gain, which is achieved by the lowest V_{MIN} and the large weighting factor for power reduction. Based on the FOM in Section VI, it shows 1.17× PPA gain with AST 5 assist in 6T-HD. Meanwhile, Fig. 13(b) shows the PPA of the 6T-HP using the SRAM assist schemes. The NBL is less effective for 6T-HP than HD due to BL IR-drop as illustrated in Fig. 4(a). The timing overhead of NBL degrades the PPA_{eff} of 6T-HP with the large weighting factor of the speed compared with the 6T-HD. The 6T-HP shows the best PPA_{eff} for the proposed DTWL (AST 6), since the DTWL provides the competitive power reduction with less timing overhead. Even if AST 5 shows the lowest V_{MIN} with the best power gain, it reduces the PPA_{eff} due to a large weighting factor of speed. The 6T-HP needs to meet the HP more over low power and small area. Therefore, the DTWL describes the best PPA gain with the optimum V_{MIN} and the minimum speed loss. The FOM of 6T-HP is induced to be 1.08× for the DTWL assist.

VIII. Conclusion

In this paper, the 10 nm FinFET 128 Mb 6T SRAM test chips are implemented to explore the various SRAM assists for the best PPA gain. The 0.040 μm^2 6T-HD and 0.049 μm^2 6T-HP bitcells are successfully demonstrated to analyze PPA with assist. Then, the proposed DTWL assist is demonstrated with 47.5 mV V_{MIN} improvement, which shows 1.08× PPA gain for the 6T-HP with small timing and area overhead. The 6T-HD SRAM explores 135 mV V_{MIN} improvement by the SBL with NBL assist showing a 1.17× PPA gain over different SRAM assists. The PPA gain FOM can be used to evaluate an SRAM assist for the different types of bitcells.

References

[1] H. Nho *et al.*, "A 32 nm high-*k* metal gate SRAM with adaptive dynamic stability enhancement for low-voltage operation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 346–347.

[2] M. E. Sinangil *et al.*, "A 28 nm 2 Mbit 6 T SRAM with highly configurable low-voltage write-ability assist implementation and capacitor-based sense-amplifier input offset compensation," *IEEE J. Solid-State Circuits*, vol. 51, no. 2, pp. 557–567, Feb. 2016.

[3] M.-F. Chang, C.-F. Chen, T.-H. Chang, C.-C. Shuai, Y.-Y. Wang, and H. Yamauchi, "A 28 nm 256 kb 6T-SRAM with 280 mV improvement in VMIN using a dual-split-control assist scheme," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 314–315.

[4] C.-H. Jan *et al.*, "A 45 nm low power system-on-chip technology with dual gate (logic and I/O) high-*k*/metal gate strained silicon transistors," in *IEDM Dig. Tech. Papers*, Dec. 2008, pp. 1–4.

[5] C.-C. Yeh *et al.*, "A low operating power FinFET transistor module featuring scaled gate stack and strain engineering for 32/28 nm SoC technology," in *IEDM Dig. Tech. Papers*, Dec. 2010, pp. 34.1.1–34.1.4.

[6] F. Arnaud *et al.*, "Technology-circuit convergence for full-SOC platform in 28 nm and beyond," in *IEDM Dig. Tech. Papers*, Dec. 2011, pp. 15.7.1–15.7.4.

[7] K. Kang, H. Jeong, Y. Yang, J. Park, K. Kim, and S.-O. Jung, "Full-swing local bitline SRAM architecture based on the 22-nm FinFET technology for low-voltage operation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 4, pp. 1342–1350, Apr. 2016.

[8] D. Jang *et al.*, "Self-heating on bulk FinFET from 14nm down to 7 nm node," in *IEDM Dig. Tech. Papers*, Dec. 2015, pp. 11.6.1–11.6.4.

[9] A. B. Sachid *et al.*, "Sub-20 nm gate length FinFET design: Can high-*k* spacers make a difference?" in *IEDM Dig. Tech. Papers*, Dec. 2008, pp. 1–4.

[10] C.-Y. Chang *et al.*, "A 25-nm gate-length FinFET transistor module for 32 nm node," in *IEDM Dig. Tech. Papers*, Dec. 2009, pp. 12.2.1–12.2.4.

[11] P. Packan *et al.*, "High performance 32nm logic technology featuring 2^{nd} generation high-*k* + metal gate transistors," in *IEDM Dig. Tech. Papers*, Dec. 2009, pp. 28.4.1–28.4.4.

[12] F. Arnaud *et al.*, "Competitive and cost effective high-*k* based 28 nm CMOS technology for low power applications," in *IEDM Dig. Tech. Papers*, Dec. 2009, pp. 28.2.1–28.2.4.

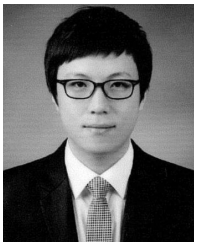
[13] C. M. Lai *et al.*, "A novel 'hybrid' high-*k*/metal gate process for 28nm high performance CMOSFETs," in *IEDM Dig. Tech. Papers*, Dec. 2009, pp. 28.3.1–28.3.4.
[14] S. K. H. Fung *et al.*, "45 nm SOI CMOS technology with 3X hole mobility enhancement and asymmetric transistor for high performance CPU application," in *IEDM Dig. Tech. Papers*, Dec. 2007, pp. 1035–1037.
[15] H. De Man, "Ambient intelligence: Gigascale dreams and nanoscale realities," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2005, pp. 29–35.
[16] H. G. Mohammadi, P.-E. Gaillardon, and G. De Micheli, "Efficient statistical parameter selection for nonlinear modeling of process/performance variation," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, to be published.
[17] M. E. Sinangil, M. Yip, M. Qazi, R. Rithe, J. Kwong, and A. P. Chandrakasan, "Design of low-voltage digital building blocks and ADCs for energy-efficient systems," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 59, no. 9, pp. 533–537, Sep. 2012.
[18] K. Zhang, E. Karl, and Y. Wang, "SRAM design in nano-scale CMOS technologies (Invited)," in *VLSI Dig. Tech. Papers*, Jun. 2012, pp. 85–86.
[19] H.-J. Cho *et al.*, "Si FinFET based 10nm technology with multi Vt gate stack for low power and high performance applications," in *2016 IEEE Symp. VLSI Technology Dig. Tech. Papers*, to be published.
[20] T. Song *et al.*, "A 14nm FinFET 128Mb 6T SRAM with VMIN-enhancement techniques for low-power applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2014, pp. 232–233.
[21] E. Karl *et al.*, "A 0.6 V 1.5 GHz 84 Mb SRAM design in 14 nm FinFET CMOS technology," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 309–311.
[22] Y.-H. Chen *et al.*, "A 16 nm 128 Mb SRAM in high-*k* metal-gate FinFET technology with write-assist circuitry for low-VMIN applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 238–240.
[23] M. Yabuuchi, Y. Tsukamoto, M. Morimoto, M. Tanaka, and K. Nii, "20 nm high-density single-port and dual-port SRAMs with wordline-voltage-adjustment system for read/write assists," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 234–235.
[24] J. Chang *et al.*, "A 20 nm 112 Mb SRAM in high-*k* metal-gate with assist circuitry for low-leakage and low-VMIN applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2013, pp. 316–318.
[25] E. Karl *et al.*, "A 4.6 GHz 162 Mb SRAM design in 22 nm tri-gate CMOS technology with integrated active VMIN-enhancing assist circuitry," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2012, pp. 230–232.
[26] H. Pilo *et al.*, "A 64 Mb SRAM in 32 nm high-*k* metal-gate SOI technology with 0.7V operation enabled by stability, write-ability and read-ability enhancements," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2011, pp. 254–256.
[27] M. Bhargava *et al.*, "Low VMIN 20 nm embedded SRAM with multi-voltage wordline control based read and write assist techniques," in *VLSI Tech. Dig. Papers*, Jun. 2014, pp. 1–2.
[28] N. N. Mojumder *et al.*, "Transistor-interconnect mobile system-on-chip co-design method for holistic battery energy minimization," in *VLSI Tech. Dig. Papers*, Jun. 2015, pp. 84–85.
[29] F. Chen and D. Gardner, "Influence of line dimensions on the resistance of Cu interconnections," *IEEE Electron Device Lett.*, vol. 19, no. 12, pp. 508–510, Dec. 1998.

Taejoong Song received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2010.

In 1997, he joined Samsung Electronics, where he developed SRAM compiler. He is currently a Principal Engineer of the Design Technology Team with System LSI, Samsung Electronics. He is in charge of the Library IP of Samsung S.LSI product. He has been published in over 12 journals and 13 conferences, and holds 25 patents. His current research interests include design technology co-optimization for high-speed and low-power in 10 and 7 nm FinFET technology.

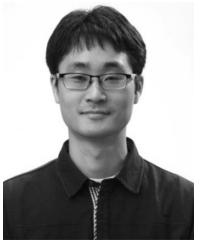
Woojin Rim was born in Seoul, South Korea, in 1979. He received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, in 2004 and 2007, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering.

From 2007 to 2010, he was with SK Hynix Semiconductor Inc., where he worked on the development of low-power DDR3 and PRAM. Since then, he has been involved with FinFET-based embedded SRAM design and SRAM compiler. Currently, he is a Senior Engineer of the Design Technology Team with Samsung Electronics, Hwaseong, South Korea. He has been published in several journals and conferences, and holds five patents. His current research interests include new circuits, process technology, device development, and design methodologies.



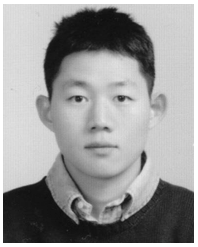
Sunghyun Park received the B.S. degree in analog and system IC design from Dongguk University, Seoul, South Korea, in 2011.

He joined Samsung Electronics, South Korea, in 2011, where he developed SRAM test vehicle with 14/10 and 7 nm FinFET process. He is currently an Engineer with the Design Technology Team of Samsung Electronics. His current research interests include 7 nm SRAM test vehicle and design of 10/7 nm embedded SRAM memory and SRAM compiler.



Yongho Kim received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2007, and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2009.

From 2009 to 2013, he was with SK Hynix Semiconductor Inc., Icheon, South Korea, where he was involved in the development of STT-MRAM and PRAM. In 2013, he joined Samsung Electronics, South Korea, where he developed SRAM. He is currently a Senior Engineer with the Design Technology Team of Samsung Electronics. His current research interests include SRAM design in 14 nm and beyond technology.



Giyong Yang received the B.S. degree in electrical engineering from Inha University, Incheon, South Korea, in 2011.

He joined Samsung Electronics, Hwaseong, South Korea, in 2011, where he developed embedded SRAM. He is currently an Engineer with the Design Technology Team of Samsung Electronics, where he is involved in low-power and high-performance standard cell development.

Hoonki Kim received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2007 and 2013, respectively.

From 2013 to 2015, he was a Post-Doctoral Associate with the University of Minnesota, Minneapolis, MN, USA. In 2015, he joined Samsung Electronics, Hwaseong, South Korea, where he was involved in the design technology co-optimization and SRAM design. His current research interests include high-speed low-power circuit design for mobile applications and embedded SRAM memory.

Sanghoon Baek received the M.S. degree in electrical and computer engineering from Sungkyunkwan University, Seoul, South Korea, in 2010.

In 2001, he joined Samsung Electronics, where he developed standard cell libraries. He is currently a Senior Engineer of the Design Technology Team with Samsung Electronics. He has been published in several journals and conferences, and holds 15 patents. His current research interests include design technology co-optimization for high-speed, low-power, and high-density standard cell development in 14 and 10 nm FinFET technology.

Jonghoon Jung received the M.S. degree in applied physics from Yonsei University, Seoul, South Korea, in 2001.

Since 2001, he has been with Samsung Electronics, where he has developed mainly memory IPs like embedded DRAM/Flash and SRAM compiler. He is currently a Principal Engineer of Design Technology Co-optimization and SRAM Test Vehicle Part with System LSI, Samsung Electronics. He has been published in several journals and conferences, and holds 21 patents. His current research interests include 14 nm FinFET SRAM macro and 10/7 nm standard cell design.

Bongjae Kwon received the M.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 1992.

In 1992, he joined Samsung Electronics, where he developed high-voltage bipolar process and ESD architecture of Library IP. He is currently a Principal Engineer in I/O library development with the Design Technology Team, Samsung Electronics.

Sungwee Cho received the B.S. degree in electrical engineering from Dongguk University, Seoul, South Korea, in 1995.

In 1995, he joined Samsung Electronics, where he developed the Standard Cell Library. He is currently a Technical Leader on Standard Cell Development with the Design Technology Team, Samsung Electronics.

Hyuntaek Jung received the B.S. and M.S. degrees in electrical engineering from Sungkyunkwan University, Seoul, South Korea, in 1995 and 1997, respectively.

In 1997, he joined Samsung Electronics, where he developed embedded DRAM. He is currently a Principal Engineer of Embedded Memory with Samsung Electronics. He holds 22 patents. His current research interests include embedded memory for OTP and STT-MRAM.

Yongjae Choo received the B.S. degree in electrical engineering from Kwangwoon University, Seoul, South Korea, in 1996.

In 1996, he joined Samsung Electronics, Hwaseong, South Korea, where he developed SRAM compiler and transferred to the library chip design field. He has been published in several journals and holds four patents. He is currently the Lead of the Chip Design Part and qualifies 20, 14, and 10 nm library test chips.

Jaeseung Choi received the B.S. degree in electrical engineering from Ajou University, Suwon, South Korea, in 1995.

From 1995 to 2002, he was with SK Hynix Semiconductor Inc., Icheon, South Korea. In 2003, he joined Samsung Electronics, where he is the Lead of the Memory Compiler Design Team, working on the design of all kinds of memory compilers such as 1-port, 2-port, ROM with high performance, low power, and small area techniques in the deep submicron region from 90 to 10 nm process technology.