Received 19 August 2019; revised 17 October 2019; accepted 19 October 2019. Date of publication 25 October 2019;
date of current version 6 December 2019. The review of this article was arranged by Editor H. T. E. Teo.

Digital Object Identifier 10.1109/JEDS.2019.2949564

A Feasibility Study on Ferroelectric Shadow SRAMs Based on Variability-Aware Design Optimization

KIYOSHI TAKEUCHI (Member, IEEE), MASAHARU KOBAYASHI (Member, IEEE), AND TOSHIRO HIRAMOTO (Member, IEEE)

Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

CORRESPONDING AUTHOR: K. TAKEUCHI (e-mail: kiyoshi.takeuchi.jp@ieee.org)

This work is based on a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

ABSTRACT A feasibility study on Ferroelectric Shadow SRAMs (FE-SRAMs) was performed using circuit simulations. To take into account design constraints set by the cell transistor variability, a simple operation margin search algorithm was proposed and used, which requires only pass/fail information from multiple transient simulations. It was found that stable dynamic recall operations can be achieved by using small enough ferroelectric capacitors, and that non-volatile write energy of well below 10 fJ/bit can be expected, adding minimal area penalty and performance degradation to the base SRAM cell. Scalability to advanced technology nodes is also anticipated. The results show that the FE-SRAM would be an ideal non-volatile memory solution for ultra-low power applications, such as sensor networks powered by energy harvesting.

INDEX TERMS Energy harvesting, ferroelectric memory, nonvolatile memory, sensor network, shadow RAM, shadow SRAM, variability.

I. INTRODUCTION

It has long been discussed that wirelessly connected sensor networks powered by energy harvesting (EH) would be a possible enabler of new smart applications [1]. EH is realized by collecting energy from various sources in environment, such as light, mechanical vibration, and radio wave. Availability of EH powered sensor devices, which do not need battery replacement, will greatly facilitate realization of large scale wireless sensor networks, since battery replacement will be a major obstacle for such applications. However, there is a drawback of EH that the power supply is unstable. It may be occasionally lost, or even only available for a limited range of time. Therefore, to avoid the loss of information during power down, some non-volatile (NV) memory function will be essential. However, embedded Flash memories [2], [3], which require very high voltage for non-volatile writing, are not suitable for achieving ultra-low power operations [4]. To overcome the limitations of Flash memories, various alternatives, such as magneto-resistive random access memory (MRAM), phase-change random access memory (PRAM), and resistive random access memory (RRAM), have been extensively studied [4]–[7]. Recently, MRAM is gaining the lead for the next generation low power embedded NV memory [8]–[10]. However, it is still not clear if these alternatives are ideal for EH applications. For example, reported MRAMs still consume relatively large energy (around 1 pJ/bit or more) for current-driven writing. Therefore, it is considered that, to realize EH sensor networks, an even lower power NV memory solution should be explored.

In this work, as a candidate for such an ultra-low power embedded NV memory, a Ferroelectric Shadow SRAM (FE-SRAM) [11]–[15] was studied. A typical FE-SRAM cell is composed of a base CMOS SRAM cell circuit and two or more additional ferroelectric (FE) capacitors. While an FE-SRAM normally functions as an ordinary volatile SRAM, the contents can be preserved during power down, by using simple "store" and "recall" operations, as detailed later.

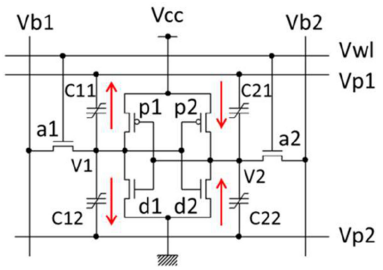

FIGURE 1. Ferroelectric shadow SRAM (FE-SRAM) cell studied in this work. Reprinted from [19].

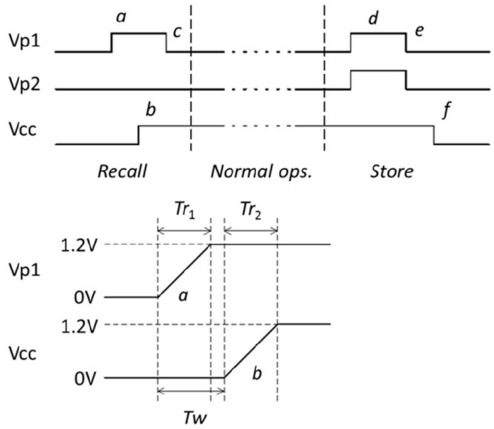

FIGURE 2. Typical FE-SRAM operation sequence, and definitions of timing parameters during recall. Reprinted from [19].

Therefore, it can pretend as if it were a true non-volatile working RAM. Unfortunately, though FE-SRAMs attracted much attention in around year 2000, their widespread use was not realized. It is considered that a major reason for this was the difficulty of scaling down conventional FE materials, due to the difficulty of etching, and poor compatibility with CMOS fabrication processes. However, now that Hafnium-based FE materials have been discovered [16]–[18], which are scalable and CMOS process friendly, the possibility of FE-SRAMs should be reconsidered.

It should be pointed out that, to assess the feasibility of FE-SRAMs, variability-aware design optimization is essential. It is well known that scaled SRAMs must be carefully designed, so that they can tolerate around $\pm 6\sigma$ (six times the standard deviation) range of transistor random variability. As for FE-SRAMs, design constraints will be also imposed by the variability, which will directly impact their usefulness and performance via the required capacitor size. Large additional capacitors will negatively impact the energy consumption and operation speed. Therefore, in this paper, a methodology for FE-SRAM variability design is first proposed. Then, the feasibility of FE-SRAMs will be discussed based on the optimization results. This paper is an extended version of the presentation in [19].

II. SIMULATION METHODS

A. CIRCUIT SIMULATION MODELS

The FE-SRAM cell examined in this work is shown in Fig. 1 [14], [15]. It comprises a normal six transistor SRAM cell, and four FE capacitors connected to the internal storage nodes V1 and V2 (6T4C cell). Two additional signal lines (plate lines) Vp1 and Vp2 are also provided. The arrows in Fig. 1 show the directions of polarization when the cell memorizes a state corresponding to V1=high and V2=low during power down. All the directions will be reversed if the opposite state is memorized. Fig. 2 shows a typical control sequence for the FE-SRAM. To wake up the FE-SRAM, a recall operation is performed, by (a) raising Vp1, and then (b) the cell power supply Vcc. This transfers the memory from the capacitors to the SRAM latch circuit. Then, normal operation follows, in which Vcc is kept high, and the cell operates as a normal volatile SRAM cell. When the cell power needs to be turned off, a store operation is performed, by (d) moving up, and (e) down the voltage of both Vp1 and Vp2. This polarizes the capacitors according to the sate of the latch, and hence, the memory is transferred from the SRAM latch to the capacitors. Then, (f) the power supply can be turned off. In this way, the memory can be preserved during power down by simple manipulation of the voltages of Vp1, Vp2, and Vcc. It is mentioned here that the polarization reversal of the ferroelectric capacitors is induced only by the store and recall operations. No reversal occurs during the normal operation, except for the very first opposite state write after recalling [15]. This will greatly relieve the requirements for endurance, which is a major reliability concern for the Hf-based ferroelectric capacitors [20].

To perform circuit simulations of the above operations, a FE capacitor model is required. In this work, a traditional Preisach model of hysteresis [21] was implemented in commercial circuit simulators, by using a Verilog-A code. Instead of adopting a simplification in [22], [23], the original form of the model was straightforwardly used as in [24]. Since a large number of simulation runs are necessary for variability design, the code was carefully optimized [25], so that sufficient speed can be obtained. As a result, the simulation time for the 6T4C cell was suppressed to 2∼3 times that of a normal 6T SRAM cell. The model parameters were extracted from measured 10 nm thick $Hf_{0.5}Zr_{0.5}O_2$ (HZO) capacitor polarization vs. voltage (P-V) curves (Fig. 3) [26], and horizontally scaled by a factor of 0.4. That is, a 4 nm FE film was assumed to match a 1.2 V power supply voltage. Switching delay model of the FE material [27] was taken into account, though it was turned off unless otherwise noted. FE-SRAM models for the simulations were formed by attaching FE capacitors of various sizes to a 6T SRAM cell net list based on a low standby power 65nm technology. The gate length and width were set to 76 nm and 120 nm, respectively, for all the six cell transistors. Gate and junction parasitic capacitances were taken into account, assuming a $1.2\mu m \times 0.48\mu m$ thin cell [28] layout (Fig. 4). Scalability to more advanced technology nodes was considered separately. The base cell was properly designed in advance to

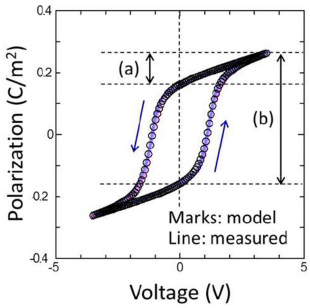

FIGURE 3. Experimental and simulated 10nm FE capacitor characteristics. For the FE-SRAM simulations, horizontal axis was scaled by 0.4. Arrows are provided to explain the difference between (a) forward and (b) reverse capacitance. Reprinted from [19].

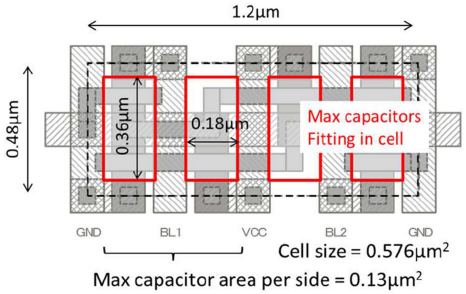

FIGURE 4. 65nm base SRAM cell. Four squares show maximum area capacitors that fit in the cell. Reprinted from [19].

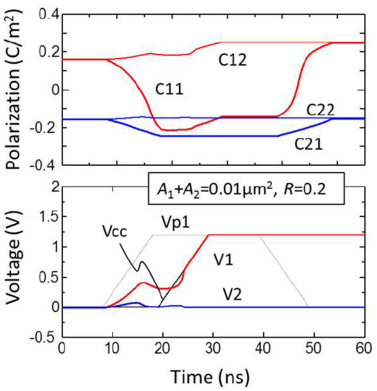

FIGURE 5. Typical recall waveforms. Reprinted from [19].

have sufficient stability margin for normal read/write operations. Only variability of cell transistors was considered here, since no data were available for the capacitor variability.

B. VARIABILITY DESIGN METHOD

Among the operations in Fig. 2, the focus of the variability-aware design will be on the recall step, since it is sensitive to the cell transistor mismatch [15]. Fig. 5 shows typical simulated voltage and polarization waveforms during the recall operation, starting from the initial polarization state shown in Fig. 1. As shown in Fig. 3, FE capacitance is larger when biased in (b) reverse direction than (a) forward direction. Therefore, V1 rises faster than V2 by the Vp1 ramping, since $C_{11}/C_{12} > C_{21}/C_{22}$; during this, C11 supplies larger

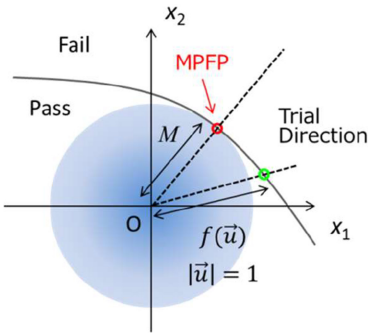

FIGURE 6. Explanation of MPFP search algorithm. Reprinted from [19].

amount of charge to the floating V1 by polarization reversal. The small V_1-V_2 is then amplified by the cell itself, which acts as a sense amplifier activated by the rising Vcc, and the SRAM cell state is recovered. Any variability of the cell transistors will result in input offset of the sense amplifier, and may cause failure of this process.

As for ordinary SRAM variability design, which is already well established, various metrics relevant to cell operations, such as static noise margin (SNM), N-curve current, and trip point voltages, can be used [29]–[31]. However, this approach cannot be easily applied to FE-SRAM recall operations, since the failure is governed by transient effects. Therefore, we propose a simple numerical method that does not rely on any such stability metrics. Fig. 6 schematically shows the principle of the method, which utilizes the concept of a Most Probable Failure Point (MPFP) [32]. Let us denote normalized threshold voltage (V_{TH}) deviations of n transistors in a circuit

$$x_i = \frac{\Delta V_{TH,i}}{\sigma_i} (i = 1, 2, \ldots, n), \tag{1}$$

where $\Delta V_{\mathrm{TH,i}}$ and σ_{i} are the V_{TH} deviation from the mean, and V_{TH} standard deviation, respectively, for the i-th transistor. The number of transistors is not fixed to six for generality, since applicability of this method is not limited to 6T SRAMs. Since random V_{TH} variation is usually normally and independently distributed, x_{i} will follow the standard normal distribution $N(0,1)$, and the joint probability density function p of the statistical variables x_{i} takes on a simple form

$$p = \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{2}}. \tag{2}$$

Let us also consider a space that is spanned by x_{i}. Fig. 6 shows such a space for $n = 2$, for graphical explanation. In this space, an equi-probability surface of (2) is given by a hypersphere centered at the origin (indicated by a shaded circle). Also, a hypersurface that determines the pass/fail boundary of the circuit operation is believed to exist in this space. The operation margin can be approximately determined by finding the closest point of the latter surface

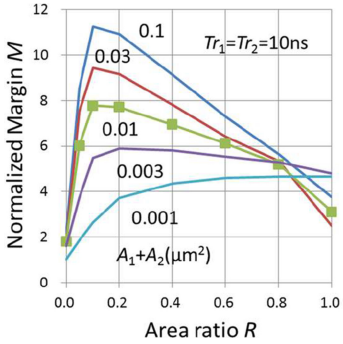

FIGURE 7. Normalized operation margin *M* vs. total area *A* and area ratio *R*. Reprinted from [19].

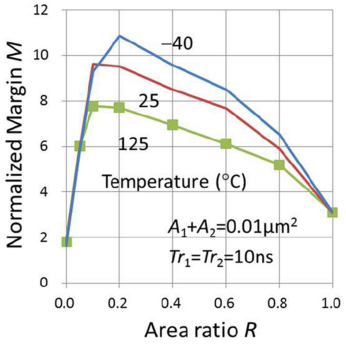

FIGURE 8. Temperature dependence of *M* vs. *R*.

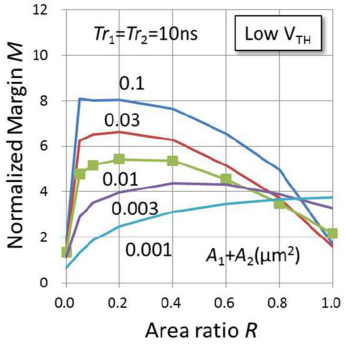

FIGURE 9. *M* vs. *A* and *R* using lower threshold voltage transistors.

to the origin, which is the MPFP; the distance between the origin and MPFP equals the normalized margin M.

The problem to be solved can be rephrased as follows: find a direction vector $u(|u| = 1)$ that minimizes a target function $f(u)$, where $f(u)$ is the length of a line segment as defined in Fig. 6. If $f(u)$ is given, this reduces to a simple non-linear optimization problem. However, here, $f(u)$ is unknown. Therefore, we propose to combine non-linear optimization with simple binary search. In the proposed algorithm, the unknown function $f(u)$ is calculated from multiple pass/fail transient simulation results, each time the trial direction u is updated by the search algorithm, by performing a binary search along the radial direction (i.e., u direction). This algorithm generates a continuous metric $f(u)$ numerically by itself. Therefore, there is no need to prepare any special continuous metric like SNM. Since only discrete pass/fail information is required, this method is applicable to various circuits and failure modes as long as the circuits pass/fail behavior can be expressed as a function of normally distributed random variables.

III. SIMULATION RESULTS

A. RECALL STABILITY MARGIN

Using the method described above, the normalized margin M was systematically calculated. The total capacitor area per side $A = A_1 + A_2$, and the ratio $R = A_1/(A_1 + A_2)$ were varied to find out the optimized design, where A_1 and A_2 are the areas of the capacitors connected to Vp1 and Vp2, respectively. $R = 0$ and $R = 1$ cases were included to fully cover both 6T4C and 6T2C cells. Vp1 and Vcc rise times (Tr_1 and Tr_2 defined in Fig. 2) were also varied as parameters. It was assumed that the word line Vwl is fixed to zero during the recall operations. Therefore, only the pull-up (p1, p2) and pull-down (d1, d2) transistor variability was considered (i.e., $n = 4$; the MPFP search was done in the four-dimensional space). Fig. 7 shows calculated margin M for $Tr_1 = Tr_2 = 10$ ns at a temperature of 125 °C, varying both A and R. It can be seen that, by using an area of only $A = 0.01\mu m^2$, combined with appropriate R values, nearly 8σ margin can be achieved. This is much smaller than the maximum area that can be accommodated in the $0.576\mu m^2$ cell ($A = 0.13\mu m^2$, see Fig. 4), and could even fit in a contact hole. Fig. 8 shows the temperature dependence. M is degraded by raising the temperature. Therefore, in this subsection, worst case 125 °C results for M are shown except Fig. 8. Fig. 9 shows M obtained by replacing the cell transistors used in Fig. 8 ($V_{TH} \sim 0.5$V at 25 °C) by those with lower V_{TH} (~0.4V). The other conditions are the same as Fig. 7. In reality, lowering V_{TH} of bulk transistors will improve the transistor variability because of reduced substrate doping, and hence, reduced random dopant fluctuation. However, for obtaining Fig. 9, the variability of the transistors was kept the same, to decouple the effects of variability change. The results show that higher V_{TH} is favorable for increasing M.

Fig. 10 shows dependence of M on Vp1 rise time Tr_1. $A = 0.01\mu m^2$. M is degraded if the Vp1 ramping speed is reduced. This shows that Vp1 and Vp2 waveforms must be carefully controlled to achieve stable FE-SRAM operations. In this paper, $Tr_1 = 10$ns was selected as a default. This is because this speed can be realized by using small width driver transistors, and would be a reasonable choice for balancing the area penalty and performance. Fig. 11 shows the effects of ferroelectric material delay on M. Here, the delay model of the Preisach hysteresis [24] was turned on, setting the characteristic delay time τ to a constant of 10ns. Tr_1 is set to 1ns, to be sufficiently shorter than τ. Comparing Figs. 7, 10 and 11, it can be seen that the material delay will degrade M, similarly to decreasing Vp1 rise time. This

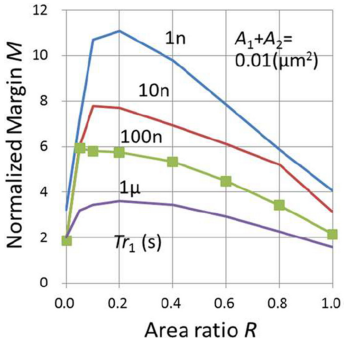

FIGURE 10. Rise time dependence of M vs. R. $Tr_2 = \min(Tr_1, 10\text{ns})$, $Tw = 1.1 \times Tr_1$. Reprinted from [19].

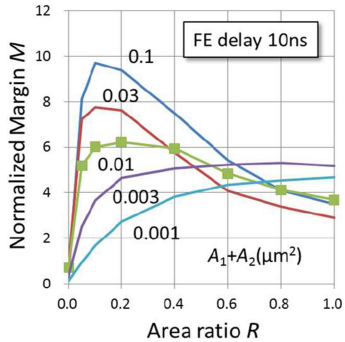

FIGURE 11. M vs. R using slow FE capacitors. $Tr_1 = 1\text{ns}$, $Tr_2 = 10\text{ns}$, $Tw = 11\text{ns}$.

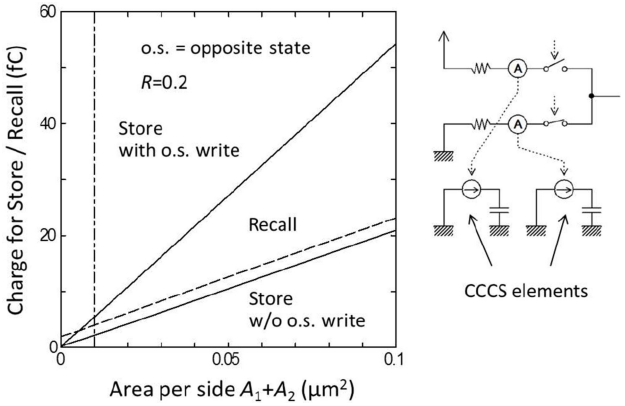

FIGURE 12. Cell level charge consumption for store and recall vs. A. Also shown is driver circuit used for evaluating the charge flow.

shows that the relatively slow switching reported for Hf-based materials [33], [34] is a concern, and that development of a fast switching FE material is very important for realizing FE-SRAMs.

B. ENERGY AND DELAY

Next, cell level energy consumption and delay of the FE-SRAM cells were evaluated, varying the FE capacitor sizes. Fig. 12 shows charge consumption necessary for the store and recall operations. The charge can be translated into energy by multiplying it by 1.2 V. The vertical straight line is provided to indicate the condition of $A = 0.01\mu\text{m}^2$

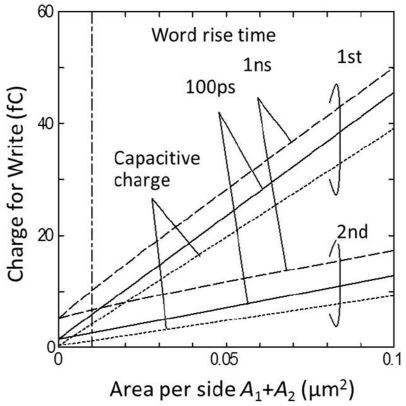

FIGURE 13. Cell level charge consumption for SRAM opposite state write vs. A.

addressed above. To obtain these results, the voltages of Vcc, Vp1, and Vp2 were driven according to the sequence in Fig. 2, using a simple driver circuit with current integration function shown in Fig. 12. The word line Vwl was kept grounded. It should be mentioned here that, as pointed out in [14], history effects exist in FE-SRAMs. The store energy depends on whether an FE-SRAM cell has experienced an opposite state write or not before the store operation, after the last recall operation. More energy is consumed in the former case, since polarization reversal of the FE capacitors is involved. In any case, if $A = 0.01\mu\text{m}^2$, the energy required for the store and recall operations is impressively low 6 fJ/bit or less. This is more than two orders smaller than the cell level write energy of a typical embedded MRAM cell.

During the normal SRAM operations, the FE capacitors will negatively impact the performance, primarily by increasing the energy consumption and delay of write operations. Fig. 13 shows the charge consumption for opposite state write. Here, Vwl was driven by a piece-wise linear functions, and the charge flow through the other terminals (mostly via Vcc and the bit lines Vb1 and Vb2) was obtained. Again, there is the history effect. Considering that SRAM cells usually consume energy on the order of several hundreds of fJ/bit or more in total for volatile read and write [35], [36] (this includes activation of long bit lines and word lines, and peripheral circuitry), the increment of the less than 5 fJ/bit at $A = 0.01\mu\text{m}^2$ is considered negligible. Fig. 14 shows internal opposite state write delay. The delay is defined as the period between the time at which Vwl reaches 0.6 V and that at which the internal node voltage switching from low to high reaches 90 % of 1.2 V. The delay increment at $A = 0.01\mu\text{m}^2$ is on the order of 0.1 ns. This would be also negligible for ultra low power applications, where relatively slow clock frequencies around 100 MHz or less are used.

The above results show that the negative effects of the FE capacitors are approximately linearly dependent on their size and that use of as small as possible FE capacitors while securing variability margin is important for achieving good FE-SRAM performance. The results also show

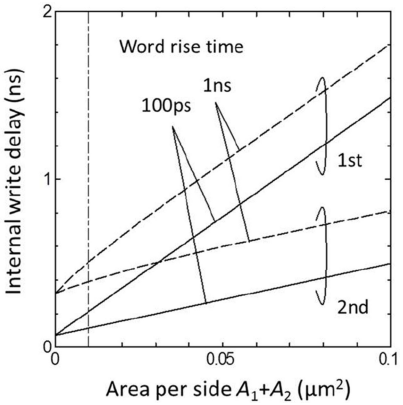

FIGURE 14. Internal opposite state write delay vs. *A*.

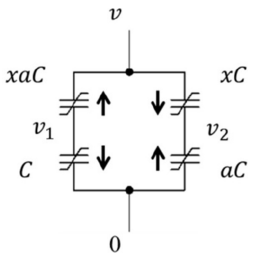

FIGURE 15. Capacitor network model for modeling FE-SRAM recall operation. Reprinted from [19].

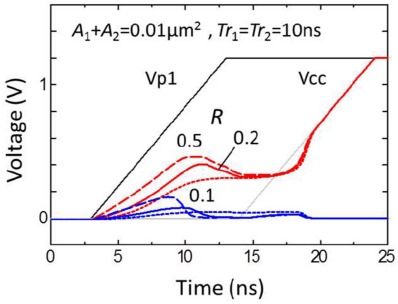

FIGURE 16. *R* dependence of recall voltage waveforms. Reprinted from [19].

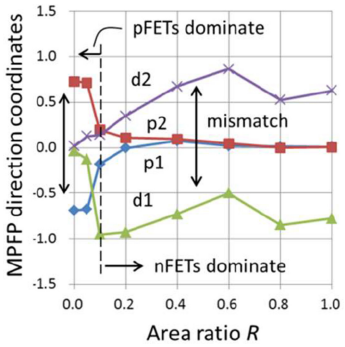

FIGURE 17. *R* dependence of MPFP directions. $A_1 + A_2 = 0.01\mu m^2$, $Tr_1 = Tr_2 = 10$ns. Reprinted from [19].

that, by properly sizing the FE capacitors, extremely small store/recall energy and almost negligible impact on normal SRAM operations can be expected.

IV. DISCUSSION

A. INTERPRETATION OF RESULTS

In this subsection, the physics behind the simulation results will be discussed. An interesting point that was found by the simulations is that M tends to be maximized when R is much less than 0.5, unless A is very small. This is contrary to what is expected from a simple capacitor network model shown in Fig. 15. Here, the ferroelectric capacitors are represented by equivalent linear capacitors for simplification. If no net charge is present in the floating nodes, the node voltages v_1 and v_2 are given by

$$\frac{v_1}{v} = \frac{ax}{1+ax}, \frac{v_2}{v} = \frac{x}{a+x}, \tag{3}$$

where $a>1$ is reverse-over-forward capacitance ratio, and x is top-over-bottom capacitor area ratio (i.e., A_1/A_2). From (3), it can be easily confirmed that v_1-v_2 is maximized when $x = 1$. Therefore, if only the capacitance matching is considered, $R = 0.5$ (i.e., $A_1 = A_2$) should be the optimum choice. To find out the reason for this discrepancy, V1 and V2 waveforms during Vp1 ramping were examined (Fig. 16). As R increases from 0.1 to 0.5, both V1 and V2 are raised higher due to the increased coupling to Vp1. However, for the cases of $R = 0.2$ and 0.5, turn-around of V1 and V2 occurs during the Vp1 ramping. This pseudo-latching behavior is caused by the leakage modulation of the cross-coupled transistors. Fig. 16 shows that, if R is too large, a ceiling is hit which is set by the leakage, and the charge induced by the FE capacitors is wasted. This shows that transistor leakage is playing a major role, in addition to the capacitive effects.

The above interpretation can be further supported by the behavior of MPFP. Fig. 17 shows direction vector (i.e., u in Fig. 6) coordinates of MPFPs vs. R. This figure indicates that, when $R = 0$, failures caused by the combination of low p1 V_{TH} (weak p1) and high p2 V_{TH} (strong p2) are most likely to occur. That is, the pFET mismatch will dominate the failure. However, when $R>0.1$, on the contrary, nFET mismatch will dominate. This can be understood by considering the fact that higher V1 and V2 voltages make nFETs more conductive, while pFETs less conductive, through the gate voltage modulation. While R is small, V1 and V2 are kept low, and pFET leakage dominates. As R increases, V1 and V2 rises, and nFET leakage dominates. As for the very small $A = 0.001\mu m^2$ case in Fig. 7, the optimum R becomes larger than 0.5. This change of the behavior would be explained by the fact that the capacitances of the capacitors and cell transistors become comparable. The storage node (V1 or V2) capacitance to ground of the base SRAM cell during Vp1 ramping, where all the terminals except Vp1 are grounded, was estimated to be 0.37fF per side. On the other hand, the capacitance of a $0.001\mu m^2$ capacitor is around 0.2fF on average (from Fig. 3, noting that the thickness was scaled by 0.4). Therefore, even if $R = 1$,

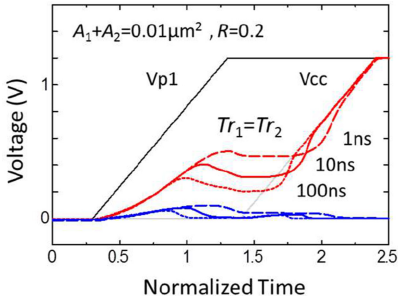

FIGURE 18. Rise time dependence of recall voltage waveforms. Reprinted from [19].

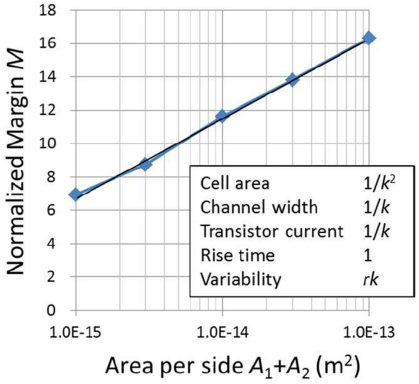

FIGURE 19. Optimized *M* vs. *A*. Variability here is lowered than other calculations by adopting the model in [38]. Reprinted from [19].

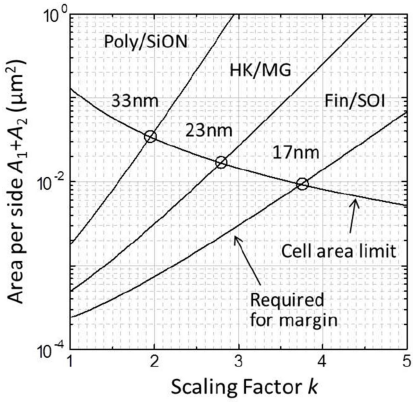

FIGURE 20. FE-SRAM scalability projection. Reprinted from [19].

the effective capacitance ratio $0.2/(0.2+0.37) = 0.35$ still remains below 0.5.

Next, the origin of the rise time dependence shown in Fig. 10 is examined. The recall waveforms for different rise time values are compared in Fig. 18. Time is normalized, so that all the waveforms can be overlaid. It can be seen that the ceiling voltage increases as the ramping speed of Vp1 increases. This can be explained as follows. The ceiling voltage is determined by the competition between capacitive current and leakage current. The leakage current is actually transistor subthreshold current, and is exponentially dependent on the storage node (i.e., gate) voltages. On the other hand, the capacitive current $C \times \mathrm{d}V/\mathrm{d}t$ is proportional to the Vp1 ramping speed. Since the ceiling voltage is reached when the leakage exceeds the capacitive current, the voltage will increase logarithmically with the ramping speed. This explains why M logarithmically increases with the ramping speed in Fig. 10. The same explanation can be also applied to the logarithmic dependence of M on capacitor area A, as is seen in Fig. 7, since capacitive current is proportional to the capacitor area, as well as $\mathrm{d}V/\mathrm{d}t$. M degradation by raising temperature (Fig. 8) or lowering V_{TH} (Fig. 9) can be also explained by the increase of the leakage current.

B. SCALABILITY PROJECTION

Finally, the scalability of FE-SRAMs to beyond 65nm node was estimated. For this purpose, theoretical projection based on a simple model was adopted to obtain transparent results. The model is based on a logarithmic M vs. A dependency as shown in Fig. 19. Here, M is maximized by independently optimizing R for each $A = A_1 + A_2$ value. In obtaining Fig. 19, realistic transistor variability model parameters (i.e., Avt values [37]) used so far were replaced by those obtained by an ideal random dopant fluctuation model [38], to take into account improvement in later technology generations. As a result, M in Fig. 19 is larger than Fig. 7. Based on Fig. 19 and the physics discussed earlier, a relationship

$$M = \frac{6.8 + 2.1 \times \ln\left(\frac{kA}{0.001\mu m^2}\right)}{rk} \tag{4}$$

is assumed, where k is the scaling factor, and r is a variability factor. In the right hand side of (4), A is multiplied by k. This is because the leakage current will decrease in proportion to $1/k$ (or channel width), which is equivalent to enlarging A. The ratio r was introduced to take into account the variability improvement achieved by adopting newer technologies. It was set to unity for poly-Si/SiON, 0.67 for high-k/metal gate (HK/MG), and 0.48 for FinFET or fully depleted silicon on insulator (FDSOI) technologies. It is assumed that the gate work function of the HK/MG devices is shifted closer to Silicon mid-gap, and the channel doping concentration is reduced. For dopant-less channel FinFET or FDSOI devices, a fixed $Avt/\sqrt{2}$ value of 1 mVμm is assumed. Fig. 20 shows calculated $A = A_1 + A_2$ necessary for achieving $M = 8.0$ vs. k, and also allowable $A_1 + A_2$ vs. k set by the cell area limitation. A stricter than 6σ criteria, i.e., 8σ, is chosen here to leave some room for additional variability. As transistors are scaled down, variability increases in proportion to $k \propto 1/\sqrt{LW}$ and the necessary $A_1 + A_2$ exponentially increases. At the same time, the allowable maximum $A_1 + A_2$ decreases in proportion to $1/k^2$. Beyond the intersections of these two curves, the SRAM cell cannot provide enough area for the required capacitors, if planar capacitors are used. At the intersections, 65 nm divided by k equals 33, 23 and 17 nm for poly-Si/SiON, HK/MG and FinFET/FDSOI,

respectively. Fig. 20 shows that the FE-SRAM would be scalable to advanced technology nodes.

V. CONCLUSION

Feasibility of Hf-based FE-SRAMs was studied, taking into account variability design optimizations. For this purpose, a simple and efficient method of operation margin estimation was proposed and used. It was found that stable FE-SRAM operation is possible by using sufficiently small FE capacitors. The physics that determine the stability were also discussed in detail. The results suggest that FE-SRAMs are feasible, and scalable to advanced technology nodes, on condition that a sufficiently fast switching FE material is available.

REFERENCES

[1] A. Keshavarzi and W. Hoek, "Edge intelligence—On the challenging road to a trillion smart connected IoT devices," *IEEE Des. Test.*, vol. 36, no. 2, pp. 41–64, Apr. 2019.

[2] Y. Taito *et al.*, "A 28 nm embedded split-gate MONOS (SG-MONOS) flash macro for automotive achieving 6.4 GB/s read throughput by 200 MHz no-wait read operation and 2.0 MB/s write throughput at Tj of 170 °C," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 213–221, Jan. 2016.

[3] N. Do *et al.*, "Scaling of split-gate flash memory with 1.05V select transistor for 28 nm embedded flash technology," in *Proc. Int. Memory Workshop*, 2018, pp. 1–3.

[4] N. Derhacobian, S. C. Hollmer, N. Gilbert, and M. N. Kozicki, "Power and energy perspectives of nonvolatile memory technologies," *Proc. IEEE*, vol. 98, no. 2, pp. 283–298, Feb. 2010.

[5] A. Chen, "A review of emerging non-volatile memory (NVM) technologies and applications," *Solid-State Electron.*, vol. 125, pp. 25–38, Nov. 2016.

[6] G. W. Burr *et al.*, "Recent progress in phase-change memory technology," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 2, pp. 146–162, Jun. 2016.

[7] B. Hudec *et al.*, "3D resistive RAM cell design for high-density storage class memory—A review," *Sci. China Inf. Sci.*, vol. 59, pp. 1–21, Jun. 2016.

[8] O. Golonzka *et al.*, "MRAM as embedded non-volatile memory solution for 22FFL FinFET technology," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2018, pp. 412–415.

[9] Y. J. Song *et al.*, "Demonstration of highly manufacturable STT-MRAM embedded in 28nm logic," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2018, pp. 416–419.

[10] K. Lee *et al.*, "22-nm FD-SOI embedded MRAM technology for low-power automotive-grade-1 MCU applications," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2018, pp. 604–607.

[11] S. S. Eaton *et al.*, "A ferroelectric nonvolatile memory," in *Proc. Int. Solid-State Circuits Conf.*, 1988, pp. 130–131.

[12] T. Miwa *et al.*, "NV-SRAM: A nonvolatile SRAM with backup ferroelectric capacitors," *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 522–527, Mar. 2001.

[13] T. Miwa *et al.*, "A 512 kbit low-voltage NV-SRAM with the size of a conventional SRAM," in *Proc. Symp. VLSI Circuits*, 2001, pp. 129–132.

[14] S. Masui *et al.*, "Design and applications of ferroelectric nonvolatile SRAM and flip-flop with unlimited read/program cycles and stable recall," in *Proc. Custom Integr. Circuits Conf.*, San Jose, CA, USA, 2003, pp. 403–406.

[15] S. Masui, T. Ninomiya, M. Oura, W. Yokozeki, K. Mukaida, and S. Kawashima, "A ferroelectric memory-based secure dynamically programmable gate array," *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 715–725, May 2003.

[16] T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide thin films," *Appl. Phys. Lett.*, vol. 99, no. 10, 2011, Art. no. 102903.

[17] J. Müller *et al.*, "Ferroelectricity in simple binary ZrO_2 and HfO_2," *Nano Lett.*, vol. 12, no. 8, pp. 4318–4323, 2012.

[18] S. Mueller, S. Slesazeck, T. Mikolajick, J. Müller, P. Polakowski, and S. Flachowsky, "Next-generation ferroelectric memories based on FE-HFO_2," in *Proc. Int. Symp. Appl. Ferroelectrics*, 2015, pp. 233–236.

[19] K. Takeuchi, M. Kobayashi, and T. Hiramoto, "A feasibility study on ferroelectric shadow SRAMs using a new variability design scheme," in *Proc. Electron Devices Technol. Manuf. Conf.*, 2019, pp. 109–111.

[20] S. Mueller, J. Muller, U. Schroeder, and T. Mikolajick, "Reliability characteristics of ferroelectric Si:HfO_2 thin films for memory applications," *IEEE Trans. Device Mater. Rel.*, vol. 13, no. 1, pp. 93–97, Mar. 2013.

[21] I. D. Mayergoyz, "Mathematical models of hysteresis," *IEEE Trans. Magn.*, vol. MAG-22, no. 5, pp. 603–608, Sep. 1986.

[22] B. Jiang, P. Zurcher, R. Jones, S. Gillespie, and J. Lee, "Computationally efficient ferroelectric capacitor model for circuit simulation," in *Proc. Symp. VLSI Technol.*, 1997, pp. 141–142.

[23] S. Mueller *et al.*, "Performance investigation and optimization of Si:HfO_2 FeFETs on a 28 nm bulk technology," in *Proc. Int. Symp. Appl. Ferroelectrics*, 2013, pp. 248–251.

[24] A. T. Batric, D. J. Wouters, H. E. Maes, J. T. Rickes, and R. M. Waser, "Preisach model for the simulation of ferroelectric capacitors," *J. Appl. Phys.*, vol. 89, no. 6, pp. 3420–3425, 2001.

[25] C. C. McAndrew *et al.*, "Best practices for compact modeling in Verilog-A," *IEEE J. Electron Devices Soc.*, vol. 3, no. 5, pp. 383–396, Sep. 2015.

[26] M. Kobayashi, N. Ueyama, K. Jang, and T. Hiramoto, "Experimental study on polarization-limited operation speed of negative capacitance FET with ferroelectric HfO_2," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2016, pp. 314–317.

[27] C. Kuhn, H. Honigschmid, O. Kowarik, E. Gondro, and K. Hoffmann, "A dynamic ferroelectric capacitance model for circuit simulators," in *Proc. Int. Symp. Appl. Ferroelectrics*, 2000, pp. 695–698.

[28] M. Khare *et al.*, "A high performance 90nm SOI technology with $0.992\mu m^2$ 6T-SRAM cell," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2002, pp. 407–410.

[29] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 5, pp. 748–754, Oct. 1987.

[30] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," *IEEE J. Solid-State Circuits*, vol. 41, no. 11, pp. 2577–2588, Nov. 2006.

[31] Y. Tsukamoto *et al.*, "Worst-case analysis to obtain stable read/write DC margin of high density 6T-SRAM-array with local Vth variability," in *Proc. Int. Conf. Comput.-Aided Design*, 2005, pp. 398–405.

[32] D. Khalil, M. Khellah, N.-S. Kim, Y. Ismail, T. Karnik, and V. K. De, "Accurate estimation of SRAM dynamic stability," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 12, pp. 1639–1647, Dec. 2008.

[33] H. K. Yoo *et al.*, "Engineering of ferroelectric switching speed in Si doped HfO_2 for high-speed 1T-FERAM application," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 2017, pp. 481–484.

[34] X. Lyu, M. Si, X. Sun, M. A. Capano, H. Wang, and P. D. Ye, "Ferroelectric and anti-ferroelectric hafnium zirconium oxide: Scaling limit, switching speed and record high polarization density," in *Proc. Symp. VLSI Technol.*, 2019, pp. 44–45.

[35] S. Miyano *et al.*, "Highly energy-efficient SRAM with hierarchical bit line charge-sharing method using non-selected bit line charges," *IEEE J. Solid-State Circuits*, vol. 48, no. 4, pp. 924–931, Apr. 2013.

[36] J. Zhu, N. Bai, and J. Wu, "A low active and leakage power SRAM using a read and write divided and BIST programmable timing control circuit," *Microelectron. J.*, vol. 44, no. 4, pp. 283–291, 2013.

[37] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in *Proc. Int. Electron Devices Meeting*, San Francisco, CA, USA, 1998, pp. 915–918.

[38] K. Takeuchi *et al.*, "Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies," in *Proc. Int. Electron Devices Meeting*, 2007, pp. 467–470.

KIYOSHI TAKEUCHI received the B.S., M.S., and Ph.D. degrees from The University of Tokyo, Tokyo, Japan, in 1984, 1986, and 1989, respectively.

From 1989 to 2015, he was with NEC, NEC Electronics, and Renesas Electronics Corporation, where he was engaged in the research of CMOS/memory devices and circuits from various aspects, including physics, modeling and design. He is currently a Project Researcher with the Institute of Industrial Science, The University of Tokyo. His current research interests are mainly on memory and power electronic devices modeling.

MASAHARU KOBAYASHI (M'09) received the B.S. and M.S. degrees in electronics engineering from The University of Tokyo, Tokyo, Japan, in 2004 and 2006, respectively, and the Ph.D. degree in electronics engineering from Stanford University, Stanford, CA, USA, in 2010.

He joined the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, in 2010. Since 2014, he has been an Associate Professor with the Institute of Industrial Science, The University of Tokyo.

Prof. Kobayashi is the Japan Society of Applied Physics.

TOSHIRO HIRAMOTO (M'92) received the B.S., M.S., and Ph.D. degrees in electronic engineering from The University of Tokyo, Japan, in 1984, 1986, and 1989, respectively.

In 1989, he joined Device Development Center, Hitachi Ltd., Ome, Japan. In 1994, he joined the Institute of Industrial Science, The University of Tokyo, and has been a Professor since 2002. His research interests include low power CMOS device design, variability in scaled transistors, silicon nanowire transistors, and silicon single electron transistors.

Prof. Hiramoto served as the General Chair of Silicon Nanoelectronics Workshop in 2003 and the Program Chair in 1997, 1999, and 2001. He served as the Program Chair of Symposium on VLSI Technology in 2013 and the General Chair in 2015. He was the Program Chair of International Conference on Solid-State Devices and Materials in 2016. He was on Committee of IEDM from 2003 to 2009. He is a fellow of Japan Society of Applied Physics and a member of IEICE.