

# Review and future prospects of low-voltage RAM circuits

Y. Nakagome  
M. Horiguchi  
T. Kawahara  
K. Itoh

*This paper describes low-voltage random-access memory (RAM) cells and peripheral circuits for standalone and embedded RAMs, focusing on stable operation and reduced subthreshold current in standby and active modes. First, technology trends in low-voltage dynamic RAMs (DRAMs) and static RAMs (SRAMs) are reviewed and the challenges of low-voltage RAMs in terms of cell signal charge are clarified, including the necessary threshold voltage,  $V_T$ , and its variations in the MOS field-effect transistors (MOSFETs) of RAM cells and sense amplifiers, leakage currents (subthreshold current and gate-tunnel current), and speed variations resulting from design parameter variations. Second, developments in conventional RAM cells and emerging cells, such as DRAM gain cells and leakage-immune SRAM cells, are discussed from the viewpoints of cell area, operating voltage, and leakage currents of MOSFETs. Third, the concepts proposed to date to reduce subthreshold current and the advantages of RAMs with respect to reducing the subthreshold current are summarized, including their applications to RAM circuits to reduce the current in standby and active modes, exemplified by DRAMs. After this, design issues in other peripheral circuits, such as sense amplifiers and low-voltage supporting circuits, are discussed, as are power management to suppress speed variations and reduce the power of power-aware systems, and testing. Finally, future prospects based on the above discussion are examined.*

## 1. Introduction

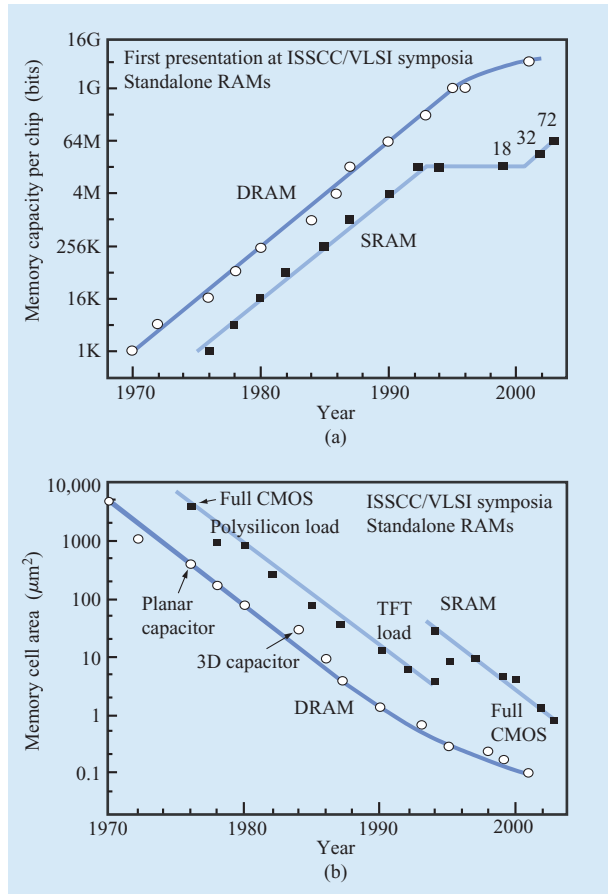
Standalone and embedded random-access memories (RAMs) have evolved rapidly, and their high density, low power, and low cost have contributed to improving the affordability and performance of electronic systems such as computers, communication systems, and consumer products. In research and development, the density of standalone RAMs has reached the 4-Gb level for dynamic RAMs (DRAMs) [1] and 72-Mb for static RAMs (SRAMs) [2, 3], along with a reduced RAM cell area, as shown in **Figure 1** [4].

In embedded RAMs (e-RAMs), recent developments have focused on high speed under low voltages,

exemplified by the 1.5-V, 300-MHz, 16-Mb DRAM macro [5] and the 1.5-V, 1-GHz, 24-Mb L3-SRAM cache [6]. Device miniaturization and the rapidly growing demand for mobile or power-aware systems have resulted in an urgent need to reduce power-supply voltage ( $V_{CC}$ ) (**Figure 2**). In standalone RAMs, the standard  $V_{CC}$  has been reduced to as low as 1.8 V. In e-RAMs, the voltage has been lowered even more, because it is based on that of the logic circuits in microprocessing units (MPUs) [7], reaching below 1.5 V. In particular, the need for e-RAMs to have low-voltage and small memory cells will become increasingly greater, because they are expected to occupy more than 90% of the area of systems-on-a-chip (SoCs)

©Copyright 2003 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/03/\$5.00 © 2003 IBM



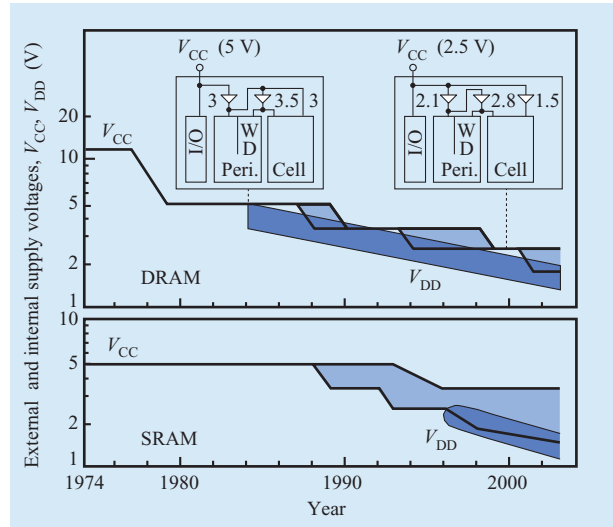
**Figure 1**

Research and development trends in DRAMs and SRAMs: (a) Memory capacity per chip. (b) Memory cell area. Data for 32-Mb [2] and 72-Mb [3] SRAMs has been added to original data [4].

[8]. Reducing the supply voltage to the region below 1 V, however, places three stringent constraints on design [4]:

- Maintaining a high signal-to-noise-ratio (S/N) for RAM cells to operate stably.
- Reducing the leakage currents (especially gate-tunnel current and subthreshold current) in MOSFETs, which increases considerably when the gate-oxide thickness ( $t_{ox}$ ) and the threshold voltage ( $V_T$ ) are reduced.
- Suppressing speed variations that become prominent at low voltages as a result of design parameter variations.

Unless these problems are solved, RAMs will never be able to operate reliably. In addition, the low-power advantage of CMOS circuits will be lost, and we can envision a scenario in which even CMOS SoCs would suffer from huge dissipations of dc power caused by



**Figure 2**

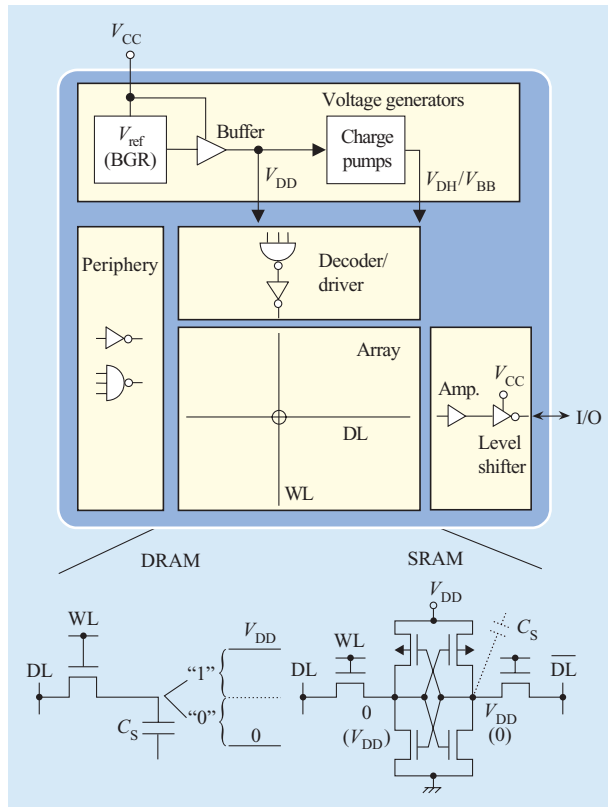
Trends in external ( $V_{CC}$ ) and internal ( $V_{DD}$ ) supply voltages for DRAMs and SRAMs. Presented at International Solid-State Circuits Conference (ISSCC) and Symposium on VLSI Circuits. Data from recent conferences has been added to original data [7].

subthreshold currents, as was the case in the recent bipolar and BiCMOS large-scale integration (LSI) eras.

In particular, reducing subthreshold current is extremely important in RAM circuit design and in random logic LSIs. To the best of our knowledge, the importance of reducing subthreshold currents in low-voltage high-speed room-temperature operation LSIs only became apparent in 1991 [9] as a result of innovative developments with 1.5-V high-speed DRAMs [10, 11]. In addition to the preceding reduction schemes through dynamic substrate control and power switches [12], other key solutions to reduce subthreshold current were proposed in the early 1990s [13–17], although these were all in the standby mode. A solution to reduce subthreshold current in the active mode was presented as early as 1993 using a hypothetical 16-Gb DRAM [18]. Although numerous attempts have subsequently been made in both RAMs and logic LSIs, the problem of reducing subthreshold current in the high-speed active mode remains unsolved, especially in random logic LSIs.

## 2. Trends and challenges with low-voltage RAMs

There are three major issues in producing low-voltage RAMs—stable RAM-cell operation, reduced leakage currents, and suppression of speed variations that are prominent at a lower voltage. However, developments toward creating a smaller cell and lower power dissipation with the simplest processes possible must also be viewed



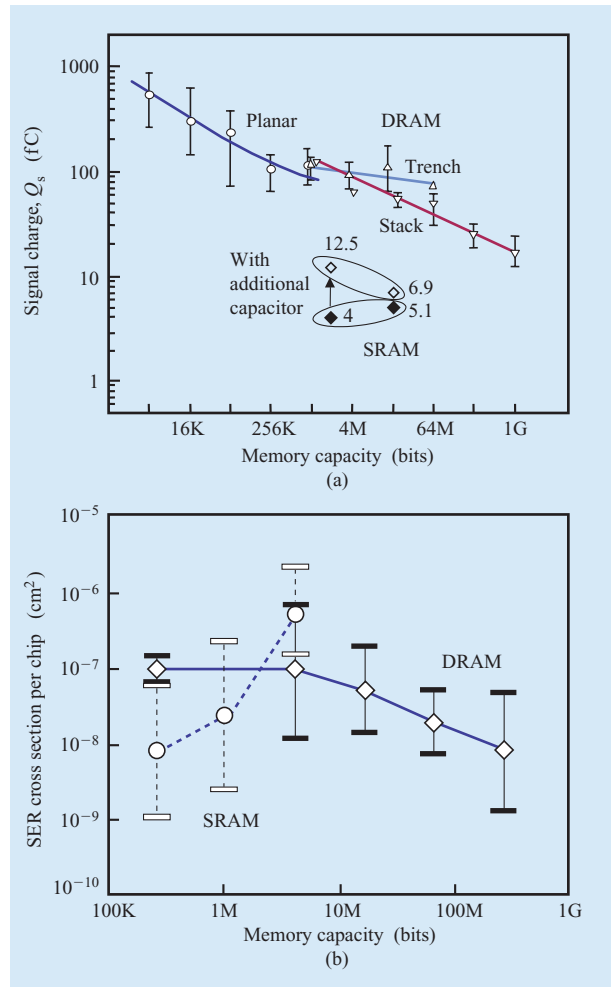
**Figure 3**

Fundamental block diagram for RAMs. A DRAM memory cell consists of one transistor and one capacitor, and an SRAM memory cell consists of six transistors.

as major concerns for RAMs, because the three issues are closely related to the degree of device miniaturization and low-voltage operation. The intention of this section is to clarify the issues common to both DRAM and SRAM technology trends. For this discussion, we have mainly assumed the standalone RAM chip shown in **Figure 3**. The chip comprises a RAM array, iterative circuit blocks such as decoders and drivers, peripheral logic circuits, I/O circuits, and on-chip voltage generators that bridge the supply-voltage gap between the memory cell array and peripheral circuits.

### Cell signal charge

The signal charge,  $Q_s$  ( $Q_s = C_s V_{DD}/2$ , where  $C_s$  is storage capacitance), has been reduced through device miniaturization and low voltage, as shown in **Figure 4(a)** [9, 19]. This reduction destabilizes DRAM-cell operations because of a smaller signal voltage on the data line (DL) in a noisy memory array and larger soft-error rates (SERs). The  $Q_s$  of SRAMs is significantly smaller than that of DRAMs by 1 to 1.5 decades. Thus, the SERs of



**Figure 4**

Trends in signal charge and soft-error immunity of RAMs: (a) Signal charge for DRAMs and SRAMs presented at ISSCC and Symposium on VLSI Circuits. Data for 1-Gb DRAM and SRAMs has been added to that reported in [9] and [19]. (b) SER cross section for DRAMs and SRAMs [20].

SRAMs increase rapidly as a result of decreased parasitic  $C_s$  and rapid reduction in operating voltage despite spatial scaling. In contrast, the SERs of DRAMs decrease gradually with device scaling, as shown in **Figure 4(b)** [20], as a result of the intentionally increased  $C_s$  and spatial scaling that causes less collection of charges.

The  $Q_s$  is effectively reduced by the ever-increasing necessary  $V_T$ ,  $V_T$  variation, and  $V_T$  mismatch under a given  $V_{DD}$ . As shown in **Figure 5(a)**, the necessary  $V_T$  of RAM cells must be increased with greater memory capacity even under ever-lowering  $V_{DD}$ . The increase in  $V_T$  is due to specifications, where the maximum refresh time,  $t_{REFmax}$ , required of standalone DRAMs must lengthen with

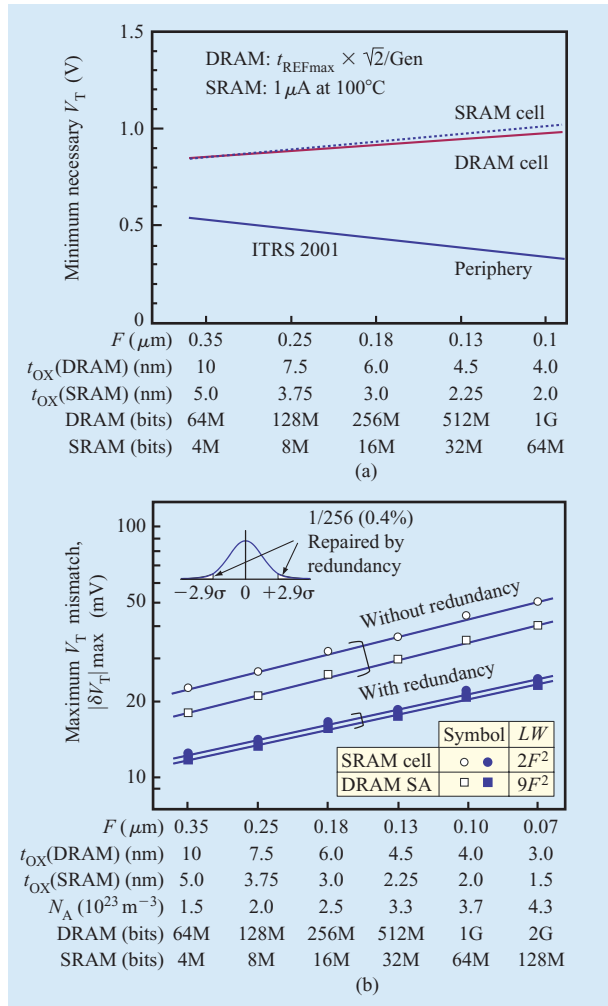


Figure 5

Threshold voltage,  $V_T$ , requirement and  $V_T$  mismatch issues in scaled RAMs: (a) Minimum necessary  $V_T$ s at room temperature to maintain the leakage charge of the DRAM cell as low as 10% of the signal charge and the retention current of SRAM chips as low as  $1 \mu\text{A}$ , both at  $100^\circ\text{C}$ , assuming  $V_T$  (extrapolated) =  $V_T$  (1 nA/ $\mu\text{m}$ ) + 0.25 V, S-factor = 120 mV/decade at  $100^\circ\text{C}$ ,  $\Delta V_T/\Delta T = -2.4$  mV/ $^\circ\text{C}$ ,  $W = 1 \mu\text{m}$  (SRAM) and  $0.25 \mu\text{m}$  (DRAM),  $Q_S = 40$  fC, and  $t_{\text{REF}} = 64$  ms for 64-Mb DRAM;  $Q_S$  is reduced by 30% and  $t_{\text{REFmax}}$  is doubled in every two generations. (b)  $V_T$  mismatch issues and possible solution (redundancy).  $V_T$  mismatch has been calculated for SRAM-cell driver MOSFETs and DRAM sense-amplifier MOSFETs, assuming their respective gate areas ( $LW$ ) are  $2F^2$  and  $9F^2$ . The depletion-layer widths  $D$  are 85, 74, 67, 59, 55, and 52 nm for  $F = 0.35, 0.25, 0.18, 0.13, 0.10, 0.07 \mu\text{m}$ , respectively. The numbers of DRAM SAs are assumed to be 65,536, 131,072, 262,144, 524,288, 524,288, 1,048,576 for 64-Mb, 128-Mb, 256-Mb, 512-Mb, 1-Gb, and 2-Gb DRAMs, respectively.

memory capacity, and the data-retention current of SRAMs in power-aware systems must almost be constant. The  $V_T$  variation slows down the half- $V_{\text{DD}}$  DRAM sensing

and reduces the available signal charge of SRAM cells. The  $V_T$  mismatch between cross-coupled/paired MOSFETs in a large number of DRAM sense amplifiers (SAs) and SRAM cells also increases with increased memory capacity and decreased device size, degrading the sensing margin of DRAM cells and the voltage margin of SRAM cells [4].

Unfortunately, even in the absence of extrinsic variations (implant nonuniformity and channel length/width variations), there is an intrinsic  $V_T$  variation that increases with device scaling as a result of random microscopic fluctuations in dopant atoms in the extremely small channel area. The standard deviation for this intrinsic random  $V_T$  variation is expressed by

$$\sigma(V_T) = \frac{q}{C_{\text{OX}}} \sqrt{\frac{N_A D}{3LW}}, \quad (1)$$

where  $q$  is the electronic charge,  $C_{\text{OX}}$  is the gate-oxide capacitance per unit area,  $N_A$  is the impurity concentration,  $D$  is the depletion layer width under the gate,  $L$  is the channel length, and  $W$  is the channel width [21]. The standard deviation of  $V_T$  mismatch (offset voltage)  $\sigma(\delta V_T)$  is  $\sqrt{2}$  times  $\sigma(V_T)$ . The maximum  $V_T$  mismatch  $|\delta V_T|_{\text{MAX}}$ , however, depends not only on the device parameters, but also on the number of MOSFETs,  $N$ , used in the chip. The ratio  $m = |\delta V_T|_{\text{MAX}}/\sigma(\delta V_T)$  increases with  $N$ , and its expected value is expressed by

$$\hat{m} = \int_0^\infty \left\{ 1 - \left[ \frac{1}{\sqrt{2\pi}} \int_{-x}^x \exp\left(-\frac{t^2}{2}\right) dt \right]^N \right\} dx. \quad (2)$$

The calculated maximum  $V_T$  mismatch in the n-MOSFETs used in DRAM SAs and SRAM cells is shown in Figure 5(b), where gate areas  $LW$  of  $9F^2$  ( $F$ : feature size) and  $2F^2$  are assumed, respectively. The mismatch is doubled with feature-size scaling from  $0.35 \mu\text{m}$  to  $0.1 \mu\text{m}$ . It should be noted that the  $\delta V_T$  in SRAM cells, as much as 50 mV in a 128-Mb SRAM, is more serious because of larger  $N$  and smaller  $LW$ . Enlarging MOSFETs to reduce the  $\delta V_T$  is fatal for a large-capacity SRAM because of increased SRAM cell area, while it can be done for DRAM SAs without substantially increasing the chip area because only one SA is placed on a pair of DLs.

One method to solve the  $V_T$ -mismatch problem of DRAM SAs is the mismatch-compensation circuit technique [22, 23], which, however, causes area and access overheads. Therefore, a column-redundancy technique is needed to eliminate a certain percentage of SAs with excessive  $\delta V_T$  to maintain the ratio  $m' = |\delta V_T|'_{\text{MAX}}/\sigma(\delta V_T)$  at a constant. Here,  $|\delta V_T|'_{\text{MAX}}$  is the maximum  $\delta V_T$  after application of a redundancy technique. For example, if the ratio of spare columns to normal columns is  $1/256$  (0.4% of array area penalty),  $|\delta V_T|'_{\text{MAX}}$  is limited to  $2.9\sigma(\delta V_T)$ . As a result, the memory capacity limitation is extended

by at least three generations, as Figure 5(b) shows. An efficient test method to detect and replace defective SAs (with excessive  $\delta V_T$ ) is also needed. On the other hand, the mismatch of SRAM cells results in random bit defects, which require quite a large number of programmable elements for storing defective addresses (three million for a 32-Mb SRAM with 128-kb spare cells). Thus, an on-chip error-checking and correcting (ECC) circuit is indispensable [24, 25].

### Leakage currents

Both subthreshold current and gate-tunnel current greatly affect the operation of RAM cells and peripheral circuits, not only in the standby mode but also in the active mode.

#### Subthreshold leakage current

In a DRAM cell, a subthreshold leakage current flowing from the cell storage node to the data line shortens the data retention time. In an SRAM, the data retention current of the cell caused by the leakage is dramatically increased, along with decreasing  $V_T$ , as Figure 6(a) shows [26]. For example, the subthreshold current of a 1-Mb SRAM array reaches as much as 10 A at  $V_T = 0$  V and 50°C, although it can be as small as 3  $\mu$ A at  $V_T = 0.65$  V, which corresponds to the maximum retention current acceptable for a standalone SRAM for cellular-phone applications. Here,  $V_T = 0$  and 0.65 V are minimum  $V_{T,s}$  corresponding to nominal  $V_{T,s}$  of 0.1 V and 0.75 V, respectively, with an assumption of a  $V_T$  variation of  $\pm 0.1$  V. Thus, the currents prevent the  $V_T$  of both DRAM and SRAM cells from scaling, as mentioned above. The leakage current in peripheral circuits, even in the active mode, also becomes huge, as exemplified in Figure 6(b) by a hypothetical 16-Gb DRAM [18]. At present, our main focus is on subthreshold current in the standby mode, because the  $V_T$  is still too high. For further reductions in  $V_T$ , however, even numerous circuits, especially the iterative circuit blocks that are inactive during the active period, will start to generate subthreshold currents, causing a huge active current in the chip.

#### Gate-tunnel leakage current

A solution to the issue of gate-tunnel leakage current is also urgently required in designing RAMs for power-aware systems because the gate-oxide thickness,  $t_{OX}$ , has been rapidly decreasing, as Figure 7 shows [27]. Recently, MPUs—and thus on-chip SRAM caches—have accelerated the trend to reduce  $t_{OX}$  at a rate of  $\times 0.175$  over the last ten years, which is almost two times faster than that for standalone DRAMs, and thus, operation of core circuits at less than 1.5 V has become popular. The  $t_{OX}$  of standard DRAMs has not been reduced so dramatically as that of MPUs (i.e., SRAMs) because of the need for stable memory-cell operations and low cost.

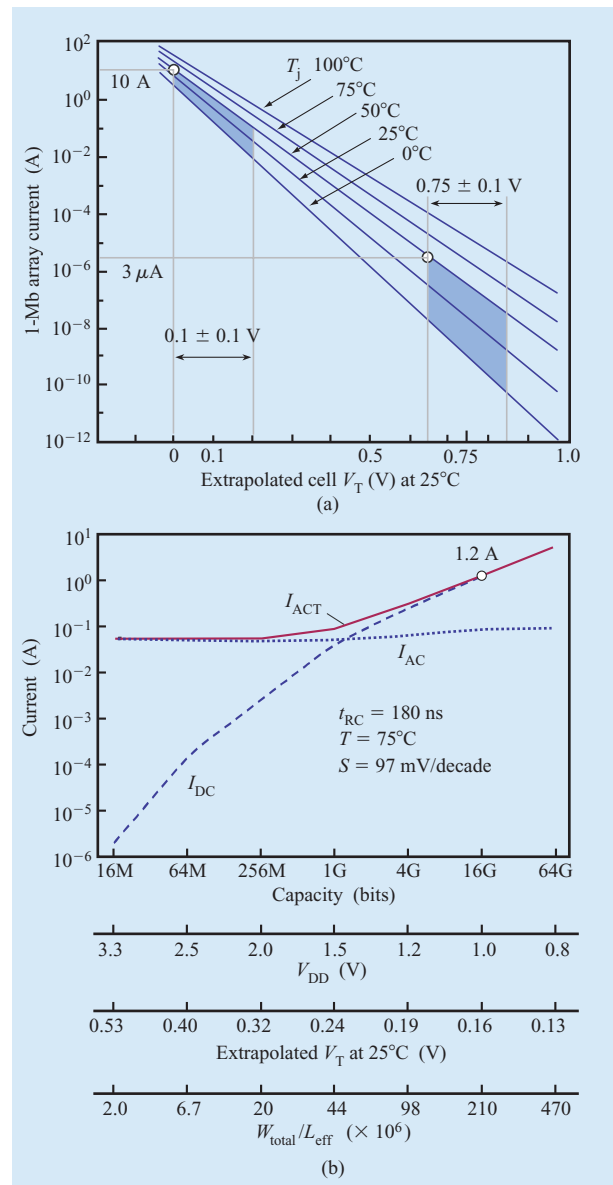


Figure 6

Leakage current issues in SRAM and DRAMs: (a) SRAM cell leakage current plotted against cell  $V_T$  for various junction temperatures,  $T_j$ . Reproduced from [26] with permission; © 1998 IEEE. (b) Trends in DRAM active current [18].

DRAM cells have needed a high operating voltage and thus, a thick- $t_{OX}$  MOSFET for stable operations with word bootstrapping, although a low-voltage—and thus a thin- $t_{OX}$ —MOSFET could be accepted for peripheral circuits. Eventually, a single thick- $t_{OX}$  MOSFET was used throughout the chip to decrease cost. Recently, however, a dual- $V_{DD}$  and dual- $t_{OX}$  device approach similar to that taken with MPUs has become popular in e-DRAMs to

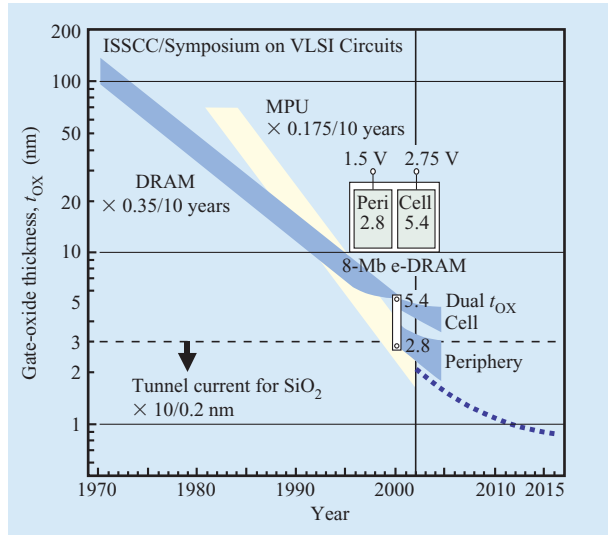


Figure 7

Trends in gate-oxide thickness,  $t_{ox}$ , for DRAMs and MPUs presented at ISSCC and Symposium on VLSI Circuits [27].

achieve higher speeds, exemplified by an 8-Mb e-DRAM with 3.7-ns access (Figure 7) [28], and a 3.3-ns-cycle 6.6-ns-access 16-Mb macro with a dual  $V_{DD}$  (1.5/2.5 V) and triple  $t_{ox}$  (1.7/2.2/5.2 nm) [5]. Even for standalone DRAMs, the dual- $t_{ox}$  approach would, in the future, be useful for high speed and low power. In this case, the thin  $t_{ox}$  of the periphery would follow the International Technology Roadmap for Semiconductors (ITRS) [29], while the thick  $t_{ox}$  of memory cells would follow a different path [Figure 5(a) and Figure 7], because it is not scalable, even if devices become increasingly miniaturized, as previously explained. Note that MPU and DRAM performances will slow down, because the pace of the  $t_{ox}$  reduction projected by the ITRS [8] will slow down. Moreover, even the ITRS projection cannot be achieved without reducing the rapidly increasing gate-tunnel current developed at a  $t_{ox}$  of less than 2–3 nm. Unfortunately, however, there have only been a limited number of circuit solutions. For example, the gate leakage current in RAM cells can be suppressed to some extent by reducing the supply voltage [25, 30]. The gate leakage current in peripheral circuits can be suppressed by shutting off the supply path by inserting a thicker- $t_{ox}$  switch [31]. The schemes can be applied only for standby mode. Since the current in the active mode must also be reduced, development of new gate-dielectric materials with low leakage and high dielectric constant appears to be the most desirable solution.

### Speed variations and other issues with peripheral circuits

It is essential to suppress speed variations of peripheral circuits because the degree of speed variation for any given variation in design parameters is increased by lowering  $V_{DD}$ , exemplified by  $\sigma(V_T)/(V_{DD} - V_T)$  [32]. Unfortunately, design parameters such as  $V_T$  increase with technology scaling, as mentioned previously. The challenge is to instantaneously raise the gate-input voltage, to reduce speed variations through stringent controls of design parameters, such as  $V_T$ , and to control  $V_T$  or compensate for  $V_T$  variation through circuit techniques. Power management is an effective way to suppress speed variations, as well as to reduce the power of power-aware systems. Testing methodology that is relevant to leakage currents is also a major area of concern.

## 3. Low-voltage RAM cells

### DRAM cells

#### One-transistor cells for standalone DRAMs

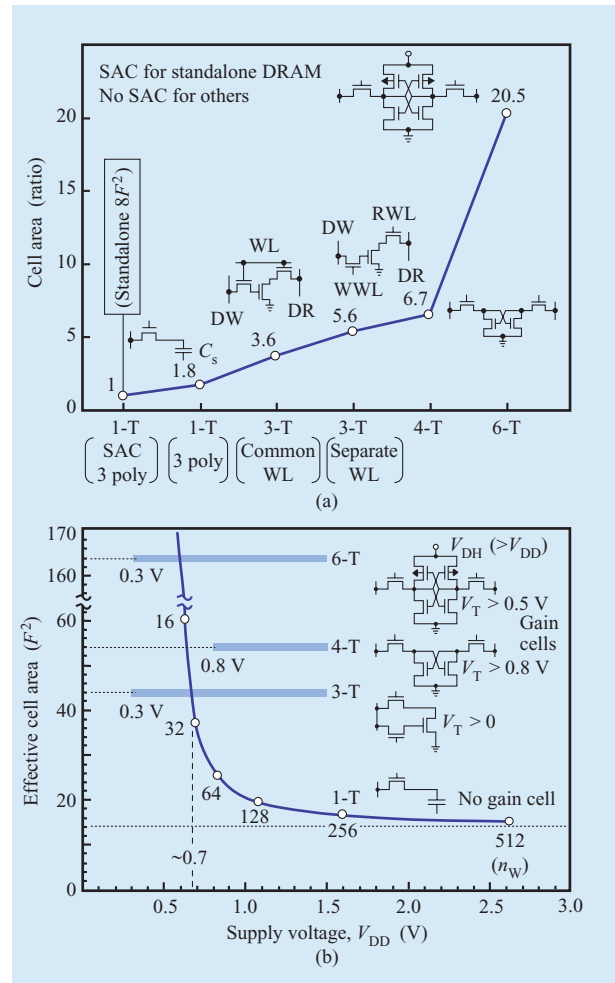
A smaller cell is the first priority in standalone DRAMs for a given cell-signal voltage ( $\approx C_s V_{DD} / 2C_D = Q_s / C_D$ , where  $C_D$  is data-line capacitance) of approximately 200 mV read out on each DL. Applying a self-aligned contact to memory cells is essential to reduce the cell area despite the speed penalty inflicted by the increased contact resistance. Leading developments of standalone DRAM cells in research and development are a  $6-4F^2$  trench-capacitor vertical-MOSFET cell [33, 34] and a  $6F^2$  stacked-capacitor open-DL cell [35]. Here, the open-DL cell necessitates a low-impedance array to suppress inherent array noises [4, 36] generated by imbalances between a pair of DLs, each of which is placed in different subarrays. For standalone DRAMs, as many memory cells as possible must be connected to each DL-pair to realize a smaller chip by reducing the overhead area at each DL-division, thus causing a larger  $C_D$ . Instead, a large signal charge,  $Q_s$ , is needed for the necessary signal voltage. Thus, a larger  $C_s$  is desirable to lower  $V_{DD}$ , which has been attained with sophisticated vertical (stacked/trench) capacitors and high dielectric constant (high- $k$ ) thin films. The subthreshold current caused by the resulting low  $V_T$  is cut by the negative word-line (NWL) scheme [4] with a  $\delta$  gate-offset during nonselected periods, as is discussed in the subsection on circuit applications in Section 4. NWL also reduces the high-level word-line voltage necessary for a full- $V_{DD}$  write operation, enabling the use of a thinner- $t_{ox}$  MOSFET for a given stress voltage [37]. Hence, low-voltage operations with a resulting small subthreshold swing (S-factor) are realized.

### One-transistor cells for e-DRAMs

The key to achieving high-performance e-DRAM is to use logic-compatible processes with a non-self-aligned cell contact and a MOS-planar capacitor and an extremely small subarray through the multi-divided DL [4]. The resultant increased cell area may be acceptable for e-DRAMs as long as it is significantly smaller than the six-transistor (6-T) full CMOS SRAM cell [7]. In addition, the resulting small  $C_s$  is accepted by the resulting small  $C_D$ , still enabling a sufficient signal voltage. Even increased SERs due to the small  $C_s$  could be solved by using an ECC [24]. The small subarray coupled with the low contact resistance of cells reduces array-relevant line delays that are major bottlenecks in the access/cycle path. Thus, DRAMs could achieve an even faster access time than SRAMs as a result of the smaller physical size of their subarrays for a given memory capacity. In addition, the small subarray, coupled with circuit techniques such as multi-bank interleaving, pipeline operation, and direct sensing [4], solves the speed problem in the row-cycle of DRAMs. A good example is the so-called 1-T SRAM\*\* [38], which incorporated a 1-T DRAM cell with a  $C_s$  smaller than 10 fF using a single polysilicon planar capacitor and an extensive multi-bank scheme with 128 banks (32 Kb in each) that can operate simultaneously. Somasekhar et al. achieved a row-access frequency higher than 300 MHz for a 0.18- $\mu\text{m}$ , 1.8-V, 2-Mb e-DRAM with a planar capacitor cell [39].

### Gain cells

Gain cells such as 3-T and 4-T cells seem to be promising when the supply voltage is reduced to less than 1 V [40]. **Figure 8(a)** compares areas of various RAM cells. The 1-T cell achieves an area of  $8F^2$  when a self-aligned contact, triple polysilicon, and vertical capacitors are used. The cell becomes larger when the contact is replaced by a non-self-aligned contact. The 3-T and 4-T DRAM cells and the 6-T SRAM cell are also shown in the figure. They do not require a special capacitor [7] and they can be fabricated by a logic-compatible process with non-self-aligned contact and single polysilicon. Obviously, in terms of the cell area and simplicity of process, the 3-T cells are attractive compared with 1-T cells and the 6-T cell. Their advantages become more prominent at a lower  $V_{DD}$ . **Figure 8(b)** compares effective cell areas for  $V_{DD}$ . Here, the effective cell area is the sum of the actual cell area and overhead area involved in the DL divisions. Note that even a high- $Q_s$  1-T cell requires more DL divisions at a lower  $V_{DD}$  to maintain the necessary signal, causing a rapid increase in the effective cell area with decreasing  $V_{DD}$  [8, 27]. The lack of gain in the 1-T cell is responsible for the increase. On the other hand, the 3-T, 4-T, and 6-T cells are all gain cells that can develop a sufficient signal voltage without increasing the number of DL divisions, even at a lower



**Figure 8**

Possible cell structures for low-voltage DRAMs [7, 27]: (a) Cell-area comparison of various cells for embedded applications. Notations for 3-T cells are write data line (DW), read data line (RW), write word line (WWL), and read word line (RWL). (b) Effective cell area including overhead area coming from the shared sense amplifier. The number of word lines ( $n_w$ ) connected to one pair of data lines has been decreased to maintain constant signal voltage of 200 mV.

$V_{DD}$ , and thus provide a fixed effective cell area that is independent of the  $V_{DD}$ . Actually, however, the  $V_{DD}$  has a lower limit for each cell. For the 3-T cell, it would be around 0.3 V, assuming a  $V_T$  for the storage MOSFET of around 0 V, an NWL scheme of  $V_{WL} = -0.5$  V for both read/write lines, and a low  $V_T$  for the read/write MOSFETs of  $V_T(r) = 0$  and  $V_T(w) = 0.3$  V. An initial stored voltage ( $V_{store}$ ) of 0.3 V for the cell, and even a decayed  $V_{store}$  of 0.1 V, can be discriminated because of the gain if an improved sensing scheme is developed. The detection of and compensation for  $V_T$  variations and

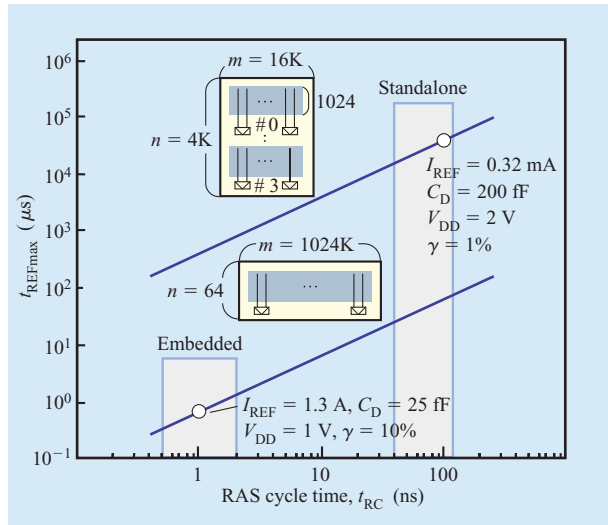


Figure 9

Maximum refresh time,  $t_{REFmax}$ , as function of RAS cycle time,  $t_{RC}$ , for 64-Mb DRAMs with logical arrays of  $4K \times 16K$  and  $64 \times 1024K$ . Device feature size of  $0.13 \mu m$  and half- $V_{DD}$  pre-charge are assumed.

an additional capacitor at the storage node would further improve stability and reliability. For the 4-T cell, it would be as high as 0.8 V, because the  $V_T$  of cross-coupled

MOSFETs must be higher than 0.8 V to ensure enough  $t_{REFmax}$ , and thus the  $V_{DD}$  must be higher than this voltage. The 6-T SRAM cell would be around 0.3 V if a raised supply voltage ( $V_{DH}$ ) (e.g., 0.5 V) were supplied from an on-chip charge pump, as explained in the next subsection. Consequently, the effective cell area of the 3-T cell would be smaller than other cells at a  $V_{DD}$  of less than 0.7 V. Note that the small polysilicon vertical-transistor 2-T  $5F^2$  cell recently proposed by Nakazato et al. [41] is another example of a gain cell, despite the small current drivability of the transistor.

In any event, in addition to the low junction temperature caused by the ultralow  $V_{DD}$ , the wide voltage margin provided by gain cells would enable a sufficient  $t_{REFmax}$ . Adjusting the potential profile of the storage node to suppress the pn-leakage current further lengthens the  $t_{REFmax}$  and preserves the refresh busy rate, even in larger-memory-capacity DRAMs [4], or it lowers the data retention current in the standby mode. Even if the  $t_{REFmax}$  were short, fast e-DRAMs, combined with a small subarray and new architectures, would allow the  $t_{REFmax}$  to be drastically shortened, as discussed in the following.

The  $t_{REFmax}$  is expressed as  $t_{REFmax} = n(t_{RC}/\gamma)$ , where  $n$  is the refresh cycle,  $t_{RC}$  is the RAS cycle time, and  $\gamma$  is the refresh busy rate, defined as  $\gamma = n(t_{RC}/t_{REFmax})$  [4]. This means that  $t_{REFmax}$  can be made smaller by reducing  $n t_{RC}$  or increasing  $\gamma$ . Figure 9 shows an example of  $t_{REFmax}$  for

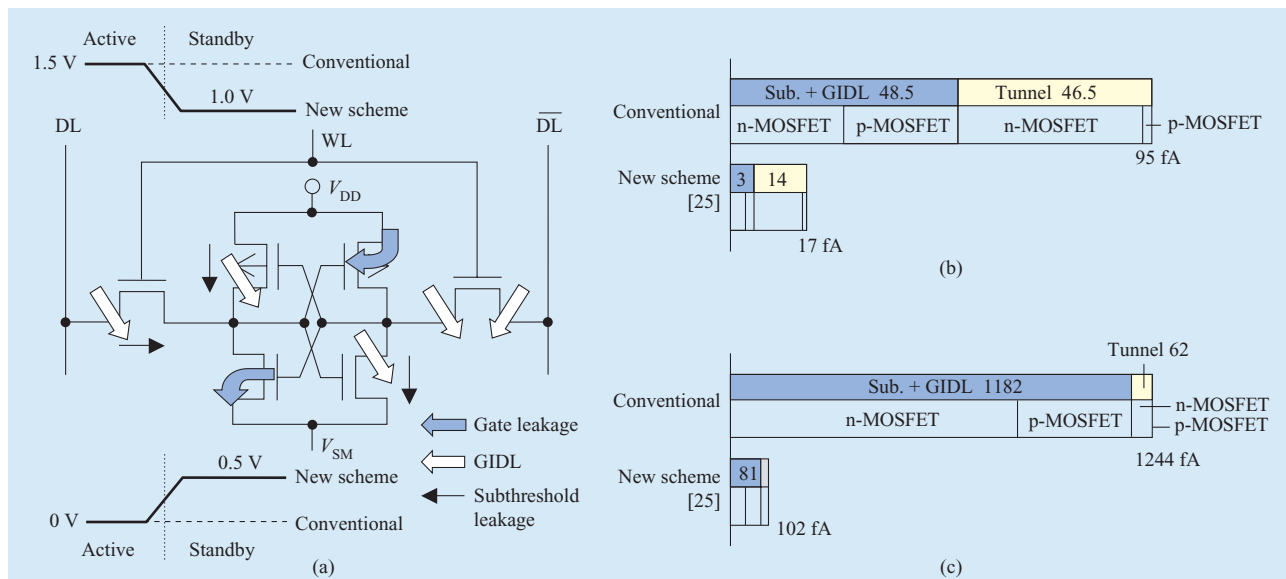


Figure 10

Leakage-current components and reduction in an SRAM cell.  $V_{DD} = 1.5 V$ ,  $V_{BB} = 0 V$ ,  $V_T$  (extrapolated) = 0.7 V (n-MOSFET), 1.0 V (p-MOSFET), and gate-oxide thickness = 3.7 nm (electrical). Reproduced from [25] with permission; © 2003 IEEE. (a) Leakage-current components; (b) measured cell-leakage currents at 25°C; (c) measured cell-leakage currents at 90°C.



a 64-Mb DRAM. There are two cases; the first is for a standalone DRAM where  $n = 4k$  (4k refresh cycles) and the second is for an e-DRAM where  $n = 64$ . Note that  $t_{REFmax}$  can be as short as  $0.64 \mu s$  for  $t_{RC} = 1 ns$ , and the refresh busy rate is 10%, while it is 40 ms for a standalone DRAM. Here, a 10% refresh busy rate may be acceptable if refreshes are hidden, as has been done in the 1-T SRAM [38]. One drawback of this scheme is to increase the refresh current ( $I_{REF}$ ) that is expressed as  $I_{REF} = M C_D V_{DD} / 2 t_{REFmax}$ , where  $M$  is the memory capacity (i.e., 64 Mb in this example).  $I_{REF}$  can increase to as high as 1.3 A in e-DRAMs, while it is as low as 0.32 mA in standalone DRAMs. However, this current may be acceptable for high-performance applications, such as the on-chip cache memories of high-performance MPUs [39].

### SRAM cells

Reducing cell area is the greatest concern in SRAMs, as is suggested by the on-chip, 3-MB, L3 cache [6]. The loadless CMOS, 4-T SRAM [42] shows promise because the cell area is only 56% of that of the 6-T cell. However, it suffers from the data-pattern problem, and it is difficult to accurately control the nonselected word-line voltage to maintain the load current. At the present time, the 6-T cell is the best, despite its large area, because it enables the use of a simple process and design made possible by the wide-voltage margin of the cell. Even in the 6-T cell, however, subthreshold currents and gate-tunnel currents as well as the gate-induced drain leakage (GIDL) increase the retention current with lowering  $V_T$  and decreasing  $t_{OX}$  [43]. Thus, this applies strict limits on how much  $V_T$  can be reduced. In addition, the soft-error issue is another concern.

To solve this problem, many driving methods and an optimal design for the cell of a small low-voltage cache have been proposed [4, 44]. Recently, a new driving scheme (Figure 10) has been proposed and applied to a 1.5-V, 27-ns access,  $6.42 \times 8.76 mm^2$ , 16-Mb SRAM [25]. The scheme, which lowers the data-line voltage from 1.5 V to 1 V and raises the ground line to 0.5 V at an active-standby mode transition, reduces the total leakage current per cell in the standby mode. At ambient temperature, the measured total current of the conventional is 95 fA. The largest component is the sum of subthreshold current and GIDL current of the n-MOSFET and p-MOSFET, although the  $V_T$ s are as large as 0.7 V and  $-1 V$ . The gate-tunnel current of the n-MOSFET is comparable to the above, despite an electrical  $t_{OX}$  as thick as 3.7 nm. The scheme greatly reduces the total current (to 17 fA). An offset source driving (discussed in the subsection on circuit applications in Section 4) by 0.5 V applied to the driver and transfer n-MOSFETs and an electric field relaxation by 0.5 V for all MOSFETs are responsible for the reduction. The reduction is more remarkable

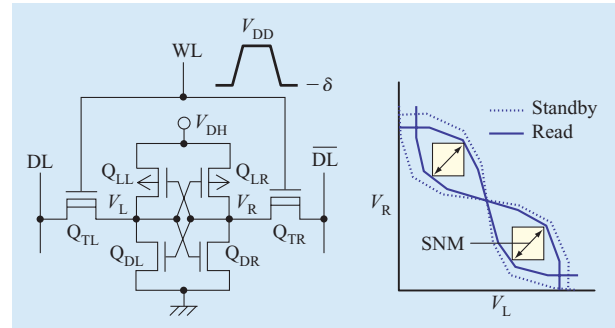
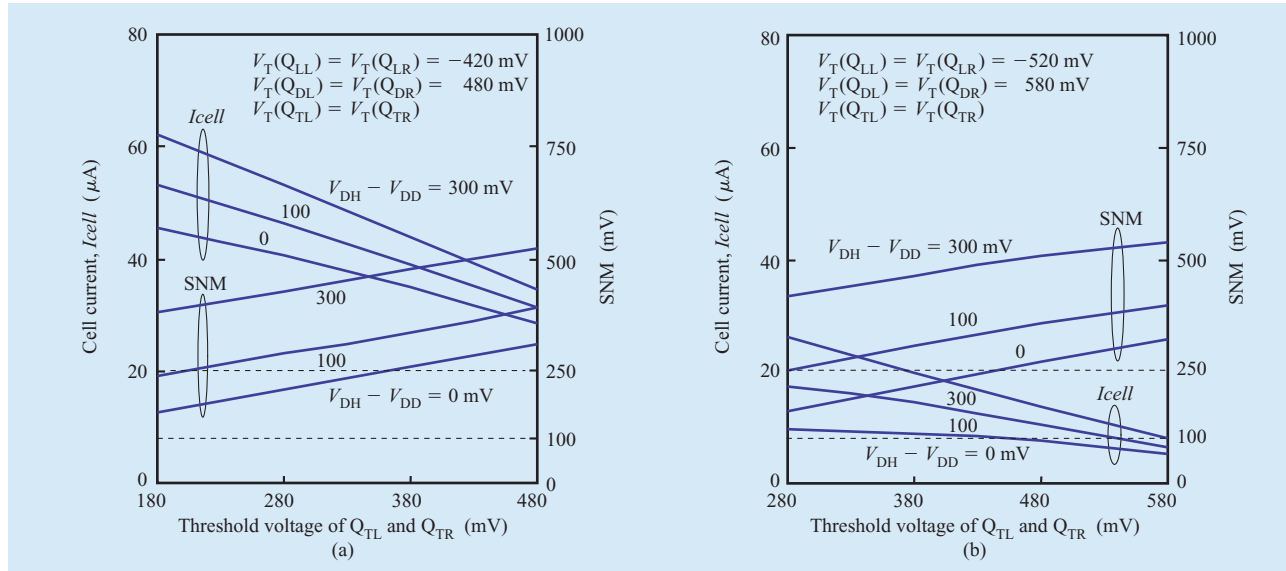


Figure 11

Improved SRAM cell [48] and static noise margin (SNM).

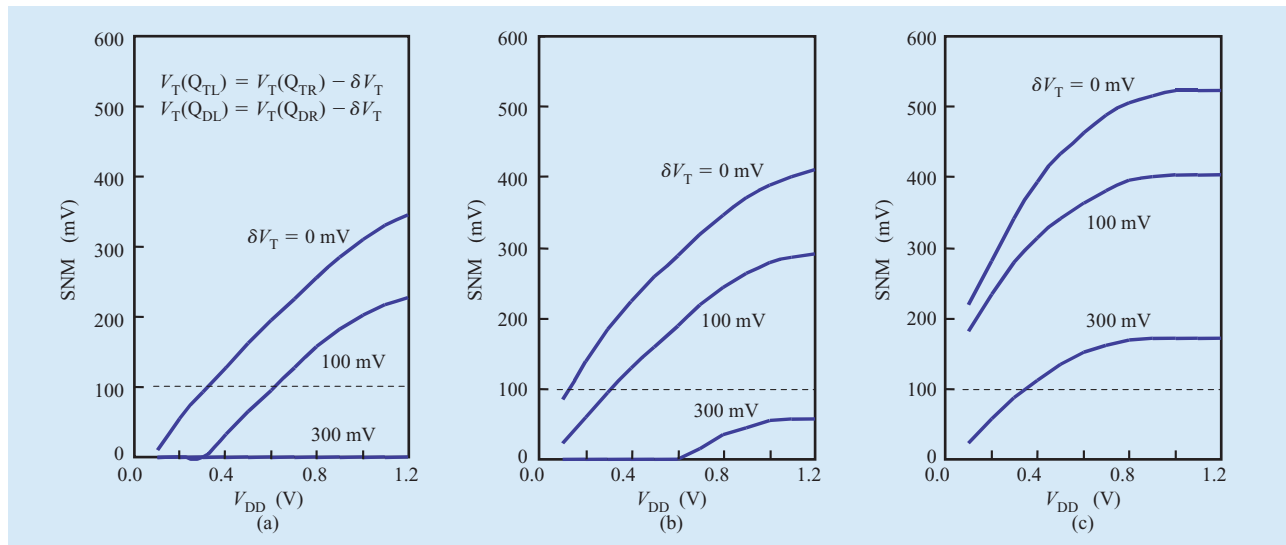
at a higher temperature. At  $90^\circ C$ , the total current of the conventional scheme is drastically increased to 1240 fA because of an increase in the subthreshold-current component. Note that GIDL current and gate-tunnel current are insensitive to temperature. The scheme reduces the total current to 102 fA. To cope with the increased SER caused by the reduced signal charge in the standby mode, an ECC was incorporated with a speed penalty of 3.2 ns and an area penalty of 9.7%, although an additional cell-capacitor can also improve the SER [Figure 5(a)] [45, 46].

Figure 11 shows another solution. The cell features a combination of a low- $V_T$  transfer MOSFET coupled with an NWL, a boosted power supply ( $V_{DH}$ ), and high- $V_T$  cross-coupled MOSFETs [47, 48]. The NWL increases cell read-current ( $I_{cell}$ ) without inducing subthreshold current in transfer MOSFETs. The high- $V_T$  MOSFETs reduce the subthreshold current. The  $V_{DH}$  increases the signal charge,  $Q_S$ , and the drivability of driver MOSFETs against the high  $V_T$  and  $V_T$  imbalance. As a result, the cell read-current and the static noise margin (SNM) are dramatically improved, as shown in Figures 12 and 13. The cell read-current increases while SNM decreases as the  $V_T$  of transfer MOSFETs decreases. However, both the current and SNM increase as the  $V_{DH}$  is raised. A usual design condition of  $I_{cell} \geq 20 \mu A$  and  $SNM \geq 100 mV$  can be realized by  $V_{DH} - V_{DD} \geq 100 mV$  at 1.0-V  $V_{DD}$  [Figure 12(a)]. Even at an 0.8-V  $V_{DD}$  and the same  $V_{DH} - V_{DD}$ , it is realized by a lower  $V_T$  of the transfer MOSFETs [Figure 12(b)]. Moreover, the cell features a strong immunity against  $V_T$  imbalance, the same as  $\delta V_T$  in the previous section. Figure 13 shows SNM calculated for the worst combination of  $V_T$  imbalance in a cell. For example, at an imbalance of 100 mV, the lower limit of  $V_{DD}$  to achieve an SNM of 100 mV is 0.6 V without boosting (i.e.,  $V_{DH} = V_{DD}$ ). However, it becomes as low as 0.3 V at  $V_{DH} - V_{DD} = 100 mV$ . There are no  $V_{DD}$



**Figure 12**

Performance of improved SRAM cell.  $W/L(Q_{LL}, Q_{LR}) = 0.18 \mu\text{m}/0.1 \mu\text{m}$ ,  $W/L(Q_{DL}, Q_{DR}) = 0.20 \mu\text{m}/0.1 \mu\text{m}$ , and  $W/L(Q_{TL}, Q_{TR}) = 0.28 \mu\text{m}/0.1 \mu\text{m}$ . All threshold voltages shown here are defined as extrapolated threshold voltage, and assumed gate-oxide thickness is 1.8 nm (optical). (a)  $V_{DD} = 1.0 \text{ V}$ ; (b)  $V_{DD} = 0.8 \text{ V}$ .



**Figure 13**

$V_T$ -imbalance immunity of improved SRAM cell.  $V_T(Q_{LL}) = V_T(Q_{LR}) = -420 \text{ mV}$ ,  $V_T(Q_{DL}) = V_T(Q_{DR}) = 480 \text{ mV}$ , and  $V_T(Q_{TL}) = V_T(Q_{TR}) = 480 \text{ mV}$ . MOSFETs are the same size as in Figure 11. All threshold voltages shown here are defined as extrapolated threshold voltage. (a) No boosting ( $V_{DH} = V_{DD}$ ); (b) boosting ( $V_{DH} - V_{DD} = 100 \text{ mV}$ ); (c) boosting ( $V_{DH} - V_{DD} = 300 \text{ mV}$ ).

limitations at  $V_{DH} - V_{DD} = 300 \text{ mV}$ . Even for an imbalance as large as 300 mV, the  $V_{DD}$  is as low as 0.35 V when  $V_{DH}$  is boosted by 300 mV. Power overhead for

generating both  $V_{DH}$  and negative word-line voltage of  $-\delta$  is negligible in the active mode. The overhead is only  $70 \mu\text{A}$ , for a total operating current of about 9 mA with

**Table 1** Concepts to create effective high- $V_T$  n-MOSFETs. The arrows indicate subthreshold leakage current ( $I_{leak}$ ).

Controlled voltage(s)		n-MOSFET	p-MOSFET
(A) $V_{GS}$ reverse biasing	(A1) $V_S$ : self-reverse biasing		
	(A2) $V_G$ : offset gate driving		
(B) $V_{BS}$ reverse biasing	(B1) $V_B$ : substrate driving		
	(B2) $V_S = V_G$ : offset source driving		
(C) $V_{DS}$ reduction			

assumptions of 128 cells per word line, a 32-b write bus, a  $V_{DH} - V_{DD}$  of 300 mV and 0.5-V  $\delta$ , and 300 MHz at a 1-V  $V_{DD}$ . In the standby mode, however, the generator current becomes larger than the total leakage current of the cell array, calling for a generator-current reduction through circuit techniques that are familiar to DRAM designers [4].

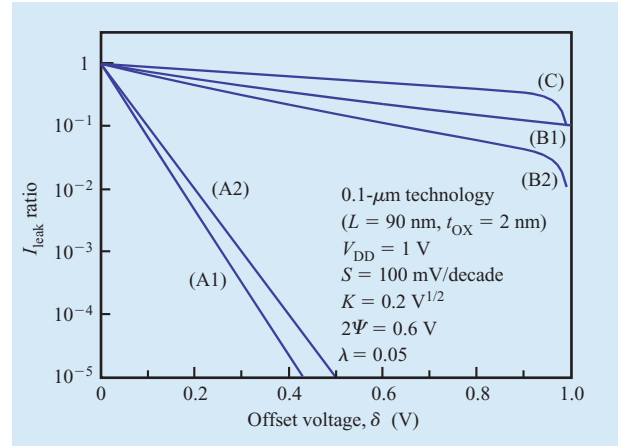
#### 4. Reduction of subthreshold current in peripheral circuits

##### Reduction scheme concepts

Increasing  $V_T$  is the best way to reduce the subthreshold current  $I_{leak}$  of a MOSFET that is expressed by

$$I_{leak} \propto \exp \left[ \pm \frac{V_{GS} - V_T - K(\sqrt{V_{BS} + 2\psi} - \sqrt{2\psi}) + \lambda V_{DS}}{S/\ln 10} \right] \times \left\{ 1 - \exp \left[ -\frac{qV_{DS}}{kT} \right] \right\}, \quad (3)$$

where plus values refer to n-MOSFETs and minus values to p-MOSFETs,  $V_T$  is the actual threshold voltage,  $S$  is the subthreshold swing,  $K$  is the body-effect coefficient, and  $\lambda$  is the drain-induced barrier lowering (DIBL) factor [49]. Here,  $q$  is the electronic charge,  $k$  is the Boltzmann constant, and  $T$  is the absolute temperature. Usually  $I_{leak}$  is reduced to 1/10 with a  $V_T$  increment of only 0.1 V (i.e.,  $S \sim 0.1$  V/decade at 100°C). The two ways of obtaining a high- $V_T$  MOSFET from a low-actual- $V_T$  MOSFET are by increasing the doping level of the MOSFET substrate and



**Figure 14**

Leakage reduction efficiency of various concepts in Table 1. Plotted using 0.1- $\mu$ m MOSFET (channel length = 90 nm, gate-oxide thickness = 2 nm) parameters.

by applying reverse biases. Thus, the selective use of the resulting high- $V_T$  MOSFETs in low-actual- $V_T$  circuits or the reverse biasing of low-actual- $V_T$  circuits decreases circuit subthreshold currents.

Although there have been many attempts to develop reverse-biasing schemes, the basic concepts can still be categorized into the three shown in **Table 1**:

- (A) Gate-source ( $V_{GS}$ ) reverse biasing.
- (B) Substrate-source ( $V_{BS}$ ) reverse biasing.
- (C) Drain-source voltage ( $V_{DS}$ ) reduction.

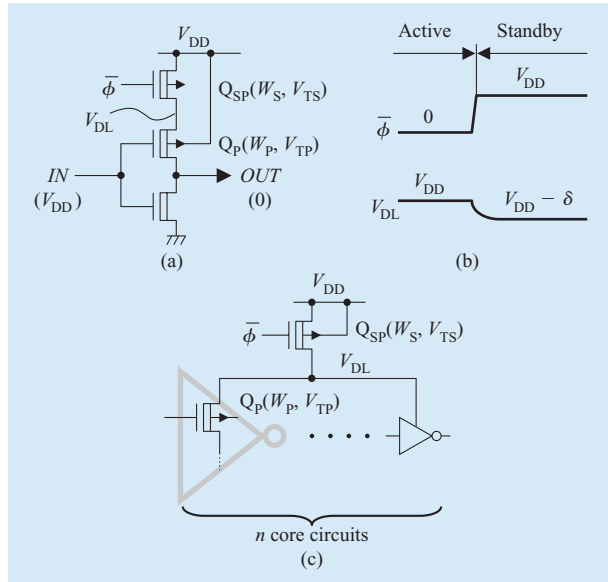
Here, the  $V_{GS}$  reverse biasing scheme can be further categorized as  $V_S$ -control with a fixed  $V_G$  (A1) [14, 15] and  $V_G$ -control with a fixed  $V_S$  (A2) [13]. The  $V_{BS}$  reverse biasing schemes can be categorized as  $V_B$ -control with a fixed  $V_S$  (B1) [12, 50] and  $V_S$ -control with a fixed  $V_B$  (B2) [51, 52].

The efficiencies for reducing leakage for offset voltage  $\delta$  are plotted in **Figure 14** using 0.1- $\mu$ m MOSFET parameters. The reduction efficiency of (A2) is the  $I_{leak}$  ratio without and with  $V_{GS}$  reverse bias:

$$r_1 = \frac{I_{leak}(V_{GS} = 0)}{I_{leak}(V_{GS} = -\delta)} = \exp \left( \frac{\delta}{S/\ln 10} \right). \quad (4)$$

This is quite large because  $\delta$  has been directly added to the low-actual  $V_T$ . The reduction efficiency of (B1) is calculated in the same manner:

$$r_2 = \exp \left[ \frac{K(\sqrt{\delta + 2\psi} - \sqrt{2\psi})}{S/\ln 10} \right]. \quad (5)$$



**Figure 15**

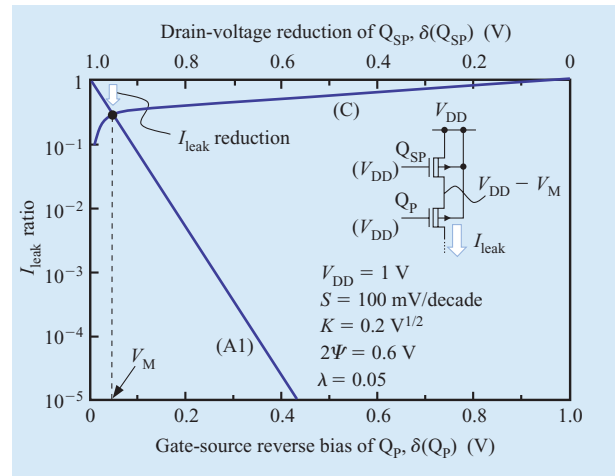
Circuits for self-reverse biasing (A1) [14, 15]: (a) Principle; (b) operating waveforms; (c) application to iterative circuits.  $W_s$  and  $W_p$  denote the respective channel widths of  $Q_{SP}$  and  $Q_P$ .  $V_{TS}$  and  $V_{TP}$  denote the respective threshold voltages of  $Q_{SP}$  and  $Q_P$ .

This is smaller than  $r_1$  because of the square-root dependence on  $\delta$  and the small  $K$ . (C) has quite a small reduction efficiency of

$$r_3 = \exp\left(\frac{\lambda \delta}{S/\ln 10}\right) \quad (6)$$

because of the small  $\lambda$ , unless  $V_{DS}$  approaches thermal voltage ( $kT/q$ ), where  $I_{leak}$  is drastically reduced as the second factor of Equation (3). Scheme (A1) has the largest reduction efficiency of  $r_1 r_2 r_3$  because all three effects are combined. (B2) has a reduction efficiency of  $r_2 r_3$ , which is larger than that of (B1) because of the additional effect of reducing  $V_{DS}$ . Note the inherently small offset voltage required to reduce the given leakage provided by scheme (A). This effectively reduces not only the subthreshold current in low-power mode, but also achieves a faster recovery time in high-speed mode, as is explained in the next subsection.

The concept involve two types of biasing, static and dynamic. The former, or so-called dual- $V_T$  scheme, is to statistically combine low- $V_T$  MOSFETs and the resulting high- $V_T$  MOSFETs in core circuits. A CMOS dual- $V_T$  scheme [53, 54] in which a low  $V_T$  is applied only to the critical path occupying a small portion of the core is quite effective in simultaneously achieving high speed and low-leakage current, although the basic scheme was proposed



**Figure 16**

Leakage reduction due to stacking MOSFETs using concepts (A1) and (C) in Table 1. The same parameters as in Figure 14 are used.

for an n-MOSFET 5-V 64-Kb DRAM [55]. A difference in  $V_T$  of 0.1 V reduces the standby subthreshold current to one-fifth its value for a single low  $V_T$ , although an excessive  $V_T$  difference might cause a race condition problem between low- and high- $V_T$  circuits. The dual- $V_T$  scheme is also applied to SRAMs [54, 56]. It was reported that a combination of dual  $V_T$  and dual  $V_{DD}$  achieved a high-speed low-power 1-V e-SRAM [56]. Another application of the dual- $V_T$  scheme is a high- $V_T$  power switch [12, 14–18] that can cut the subthreshold current of an internal low- $V_T$  core in standby mode, as described in the subsection on circuit applications. High- $V_T$  MOSFETs can easily be produced in a DRAM [57] by using the internal supply voltages that are required by DRAMs, as explained in the subsection on applications to RAMs. The high  $V_T$ , however, eventually restricts the lower limit of  $V_{DD}$  as the transconductance of the MOSFET degrades at a lower  $V_{DD}$ .

The latter changes the  $V_T$  so that it is low enough in high-speed modes, such as active mode with no reverse bias, while in low-power modes, such as standby mode, it is increased by changing bias conditions, as shown in Table 1.

### Circuit applications

This section reviews dynamic biasing schemes based on the above basic concepts, assuming circuits in which all MOSFETs have a low actual  $V_T$ .

#### Gate-source self-reverse biasing (A1)

Figure 15(a) is a circuit diagram for self-reverse biasing. It features a low- $V_T$  switch p-MOSFET  $Q_{SP}$  inserted between

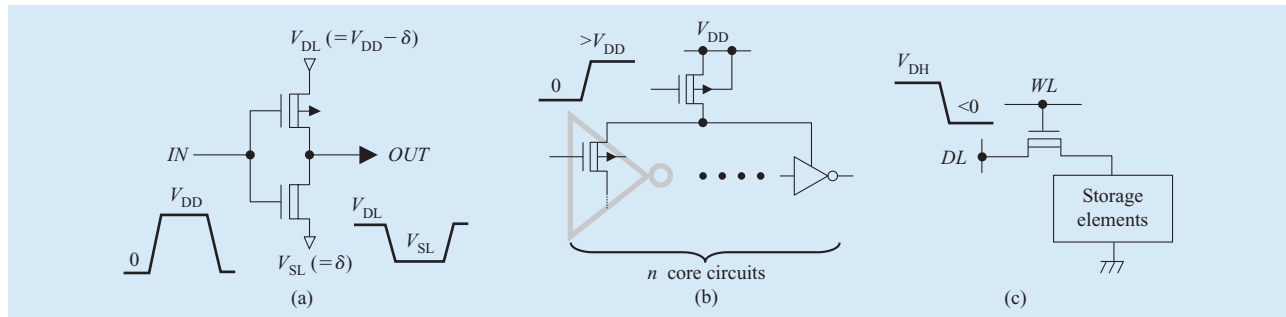


Figure 17

Circuits for offset gate driving (A1) [13]: (a) Principle; (b) application to power switch [60]; (c) application to RAM cells (negative word line) [47, 61].

the source of the MOSFET  $Q_p$  and  $V_{DD}$ . The MOSFET  $Q_{sp}$  stacked to  $Q_p$  is a kind of power switch, working as a source impedance turning on and off during respective active and standby modes. A subthreshold current flowing from  $Q_p$  when  $Q_{sp}$  and  $Q_p$  are off in the standby mode generates an offset voltage,  $\delta$ , on  $V_{DL}$  as shown in **Figure 15(b)**, automatically providing a reverse bias  $\delta$  to  $Q_p$  so that the current is eventually reduced. This biasing is a combination of  $V_{GS}$  reverse biasing,  $V_{BS}$  reverse biasing, and  $V_{DS}$  reduction, providing the primary effect to  $V_{GS}$  reverse biasing and the secondary effect to  $V_{BS}$  reverse biasing and  $V_{DS}$  reduction, as described above. The gate voltage is  $V_{DD}$ , not  $V_{DL}$ , to take advantage of the  $V_{GS}$  reverse bias. Note that no matter how large the original leak current at  $Q_p$  is, it is eventually confined to the constant current of  $Q_{sp}$  through the automatic adjustment of the offset voltage  $\delta$ . Here,  $\delta$  is expressed as  $V_{TS} - V_{TP} + S \log(W_p/W_s)$ , and the current reduction ratio is expressed as  $10^{-\delta/S}$  if secondary effects are neglected [4]. Thus, the reduction is adjustable with  $\delta$ , that is,  $V_{TS}$  and  $W_s$ . If  $V_{TS}$  is high enough, the current is completely cut off with a larger  $\delta$ , creating a perfect switch. A large  $\delta$ , however, results in slow recovery time, large charging/discharging current, and spike noise at mode transients. If  $V_{TS}$  is low enough, however,  $\delta$  becomes smaller (allowing leakage flow), causing an imperfect (leaky) switch, but the above problems are reduced. Moreover, a low- $V_T$  switch is favorable to reduce the necessary channel width of  $Q_{sp}$ , because the increased transconductance can supply the accumulated current of the logic core with a smaller channel width, especially at a lower  $V_{DD}$ . Sharing a low- $V_T$  switch through iterative circuits in RAMs [**Figure 15(c)**] is quite effective [14, 15]. Because a feature of RAM circuits is that only one of the iterative circuits is active,  $W_s$  can be comparable to

$W_p$  with little speed penalty in the active mode, while  $\delta = S/\log(nW_p/W_s)$  in the standby mode for  $V_{TS} = V_{TP}$ . Therefore, both leakage and area penalty as a result of adding  $Q_{sp}$  are negligible with increasing  $n$  (i.e.,  $\delta$ ). To be more precise, secondary effects must be taken into consideration: The substrate connection of  $Q_p$  to  $V_{DD}$  creates substrate reverse bias. The effect of reduced  $V_{DS}$  is also added if  $\delta$  is large (i.e., a small  $V_{DS}$ ).

An extreme case of  $W_s = W_p$  and  $n = 1$  is in the  $I_{leak}$  reduction of series-connected MOSFETs, the so-called *stacking effect* [58, 59]. This effect can be explained by a combination of self-reverse biasing (A1) and  $V_{DS}$  reduction (C), as **Figure 16** shows, though (C) is not used alone. The leakage current of  $Q_p$  is reduced through self-reverse biasing, while that of  $Q_{sp}$  is reduced through reducing  $V_{DS}$ . The node-voltage-lowering  $V_M$  at the connection and the  $I_{leak}$  reduction efficiency are determined by the equilibrium of the two currents and expressed by the crossing point of the two curves. Because the reduction efficiency becomes larger as the number of series MOSFETs becomes larger, the  $I_{leak}$  of NAND gates using series-connected n-MOSFETs is efficiently reduced.

#### Offset gate driving (A2)

**Figure 17(a)** shows offset gate driving, where the input voltage is “overdriven” by  $\delta$ . This is difficult to apply to random logic circuits because the logic swing of the output must be smaller than that of the input. However, it is useful to reduce  $I_{leak}$  in bus drivers [13], in power switches that have a low actual  $V_T$  (**Figure 17(b)** [60]), and in RAM cells (**Figure 17(c)** [47, 61]), as was previously explained. Offset gate driving applied to an imperfect switch reduces  $I_{leak}$  in standby, realizing an effectively perfect switch. However, the problems of a perfect switch described above arise.

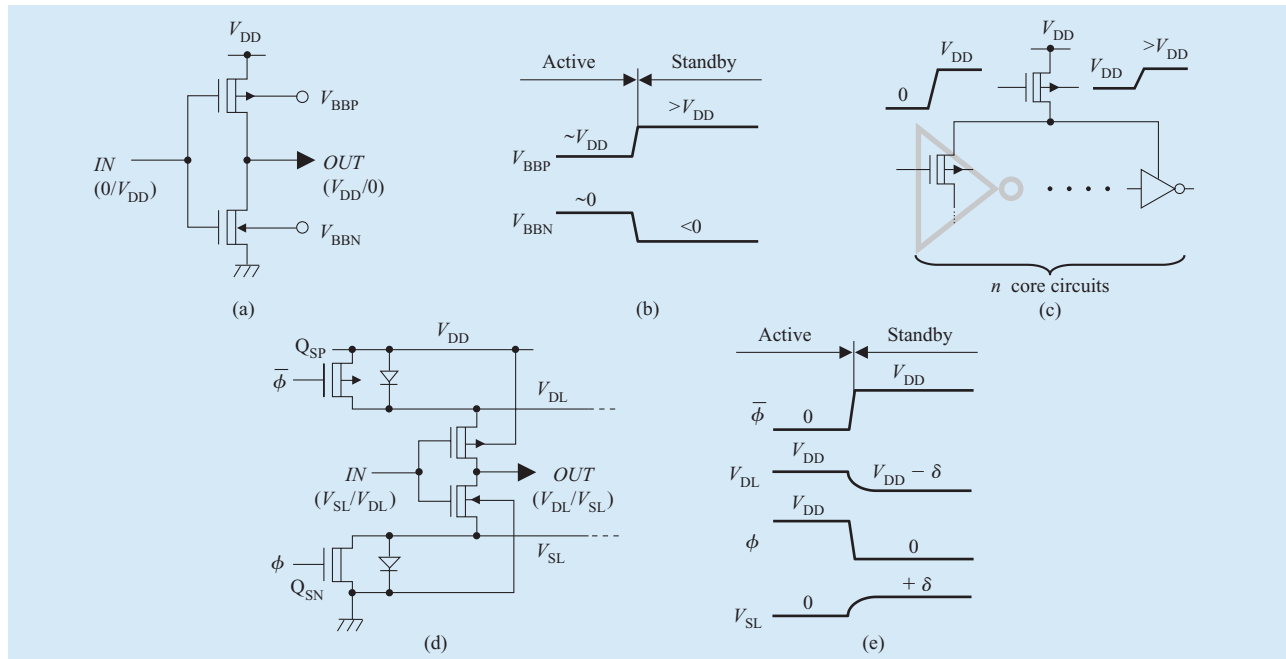


Figure 18

Circuits for substrate-source voltage ( $V_{BS}$ ) reverse biasing: (a) Substrate (well) driving (B1) [12, 50]; (b) its operating waveforms; (c) application to power switch [64]; (d) offset source driving (B2) [51, 52]; (e) its operating waveforms.

### Substrate (well) driving (B1)

Figure 18(a) shows the circuit for substrate (well) driving, where the substrate voltages of MOSFETs in core circuits change between active and standby modes [12, 50, 62, 63]. Figure 18(b) shows the operating waveforms. This scheme can also be applied to reduce  $I_{leak}$  in power switches (Figure 18(c) [64]).

### Offset source driving (B2)

Figure 18(d) [51, 52] has the circuit for offset source driving, with switches  $Q_{SP}$  and  $Q_{SN}$  inserted between the MOSFET sources and power supplies. Note that this is quite different from (A1), though both utilize source switches. The input (gate) voltage of (A1), which is the output of the previous stage, is “full swing” ( $V_{DD}$ ), while that of (B2) is not (i.e.,  $V_{DL}$  or  $V_{SL}$ ). This difference results in the large discrepancy in  $I_{leak}$  reduction efficiency, as shown in Figure 14. From this viewpoint, power switches [17] applied to logic circuits can be categorized as (B2). Another application of this scheme is to reduce  $I_{leak}$  in SRAM cells [25, 65], as was discussed earlier.

### Comparison

There is a big difference between the two schemes (A) and (B) in mode-transient time, especially recovery

(standby-to-active) time. In  $V_{GS}$  reverse biasing, the small voltage swing,  $\delta$ , enables quick recovery (several nanoseconds). In  $V_{BS}$  reverse biasing, however, it takes more than 100 ns for recovery when it is applied to a power line, because  $V_{BS}$  reverse biasing requires a large  $V_B$  swing ( $\Delta V_B$ ) or  $V_S$  swing ( $\Delta V_S$ ), which is usually more than 1.5 V for a given change in  $V_T$  ( $\Delta V_T$ ). The necessary voltage swing imposes different requirements on substrate driving (B1) and offset source driving (B2). In (B1), the necessary voltage is significantly larger than  $V_{DD}$ , which is the sum of  $V_{DD}$  and  $\Delta V_B$ . For example, existing MOSFETs with a  $0.2\text{-V}^{1/2}$ -body-effect coefficient ( $K$ ) require a  $\Delta V_B$  as large as 2.5 V to reduce the current by two decades with a  $0.2\text{-V}$   $\Delta V_T$ . A larger- $K$  MOSFET is needed to reduce the swing. However, this slows down the speed in stacked circuits, such as NAND gates. In contrast, the  $K$  value decreases with MOSFET scaling, implying that the necessary  $\Delta V_B$  will continue to increase further in the future owing to a lower  $K$ , and there will be a need for a larger  $\Delta V_T$  reflecting the low- $V_T$  era. Eventually, this will enhance short-channel effects and increase other leakage currents, such as the GIDL current [66]. A shallow reverse  $V_B$  setting, or even a forward  $V_B$  setting in active mode, is also required to effectively increase  $V_T$  in standby mode, because  $V_T$  is more sensitive to  $V_B$  [4]. However, the

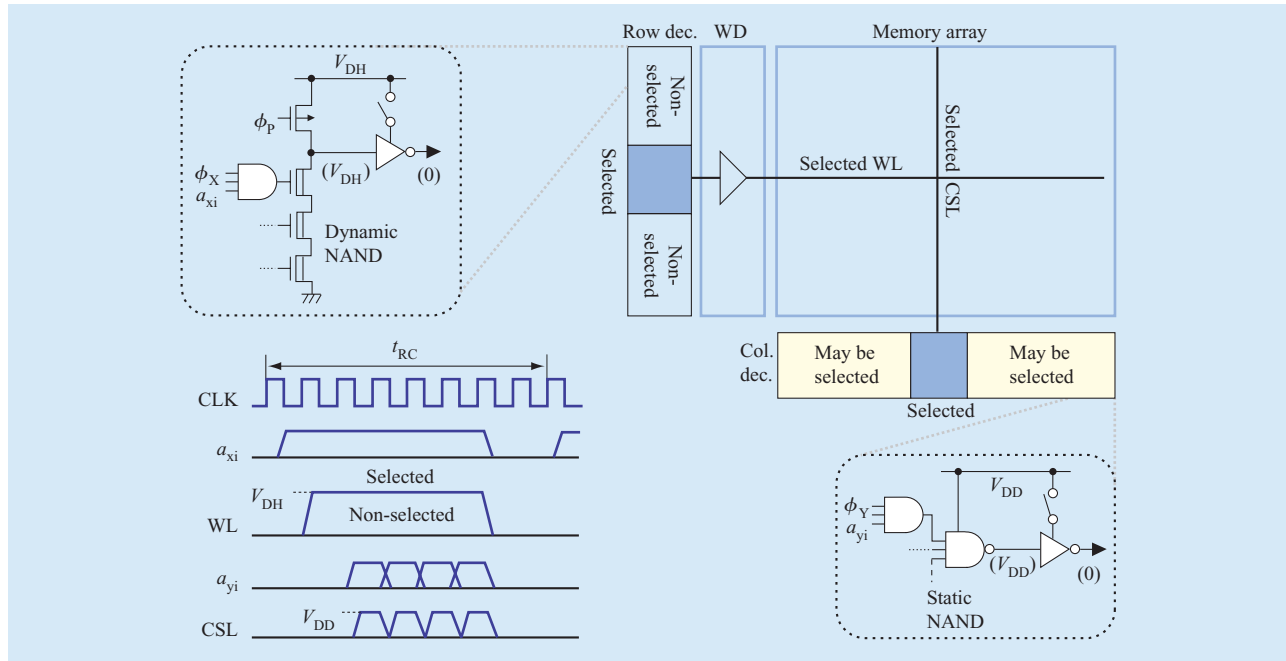


Figure 19

Features of DRAM circuits in terms of leakage reduction. WL and CSL indicate respective word line and column selection line. Each node voltage during standby mode is in parentheses.

requirements to suppress  $V_B$  noise will instead become more stringent. In fact, a connection between the substrate and source every 200  $\mu\text{m}$  [63] to reduce noise has been proposed, despite an area penalty. In addition, problems inherent in LSIs with an on-chip substrate bias ( $V_{BB}$ ) generator, which DRAM designers have experienced since the late 1970s, may occur even though  $V_{DD}$  is low. These problems include spike current and CMOS latch-up during power-on and mode transitions,  $V_{BB}$  degradation caused by increased substrate current in high-speed modes and screening tests at high stress  $V_{DD}$ , and slow recovery time as a result of poor current drivability of the on-chip charge pump.

In offset source driving (B2), the necessary voltages and voltage swing at any node are smaller than  $V_{DD}$ . This control becomes ineffective as  $V_{DD}$  is lowered owing to a smaller substrate bias. However, the problems described above accompanied by an on-chip  $V_{BB}$  generator are not expected.

The energy overhead of offset source driving (B2) through mode transitions is usually larger than that of substrate driving (B1). This is because the parasitic capacitances of source lines ( $V_{DL}$  and  $V_{SL}$ ) are larger than those of substrate lines ( $V_{BBP}$  and  $V_{BBN}$ ), though the necessary  $\delta$  is smaller, as shown in Figure 14. The parasitic capacitances of  $V_{BBP}$  and  $V_{BBN}$  consist mainly of junction capacitances

between substrate (well) and source/drain of MOSFETs, while those of  $V_{DL}$  and  $V_{SL}$  include the gate capacitances of on-state MOSFETs as well as junction capacitances. The energy overhead of self-reverse biasing (A1) is quite small because of small and self-adjusted  $\delta$ .

### Applications to RAMs

#### Features of RAMs

In the active mode, reducing leakage is extremely difficult because of the limited time to control it. In the standby mode, it is rather easy because there is sufficient time available. Fortunately, however, RAM peripheral circuits favor the reduction of subthreshold current ( $I_{\text{leak}}$ ) (Figure 19) compared with random logic gates, because of the inherent features of RAMs described in the following. These are exemplified by the modern synchronous DRAM in the figure.

#### Use of iterative circuit blocks

RAMs consist of multiple iterative circuit blocks with low activation ratios, such as row/column decoders and drivers, each of which has quite a large total-channel width involving subthreshold current. In addition, all circuits in each block, except the selected one, are inactive, even

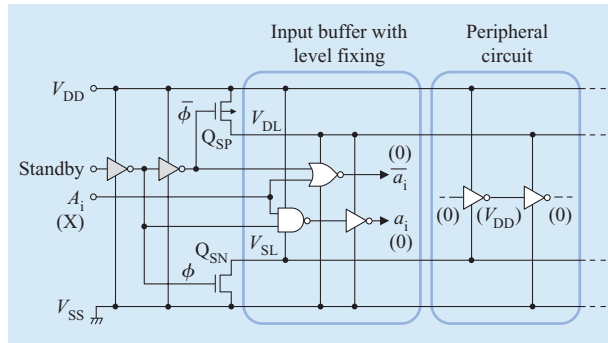


Figure 20

Method to make the internal nodes of RAMs predictable. Each node voltage during standby mode (standby signal is at high level) is in parentheses.  $Q_{SP}$ ,  $Q_{SN}$ , and solid inverters consist of high- $V_T$  MOSFETs. Other logic gates consist of low- $V_T$  MOSFETs. Reproduced from [15] with permission; © 1993 IEEE.

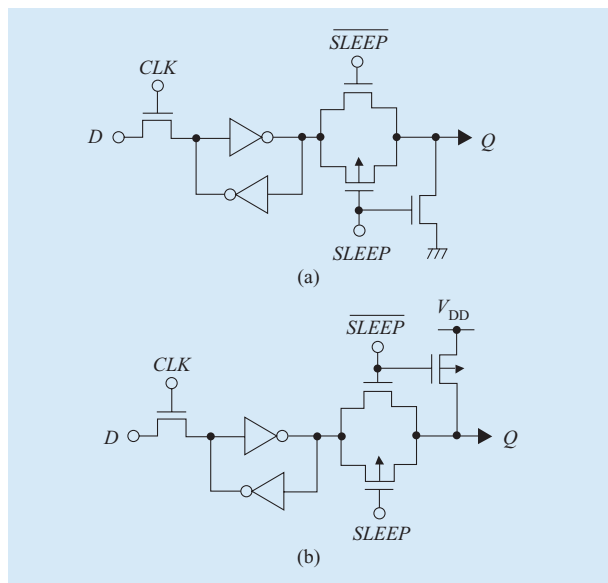


Figure 21

Latches with output fixing in sleep mode: (a) Output fixed low; (b) output fixed high in sleep mode. Reproduced from [59] with permission; © 1997 IEEE.

during the active period. This enables  $I_{leak}$  to be controlled simply and effectively with a smaller area penalty than logic LSIs, as shown in Figure 15(c).

#### Use of input-predictable logic

RAMs are composed of input-predictable circuits, allowing circuit designers to predict all node voltages in the chip and to prepare the most effective subthreshold-current

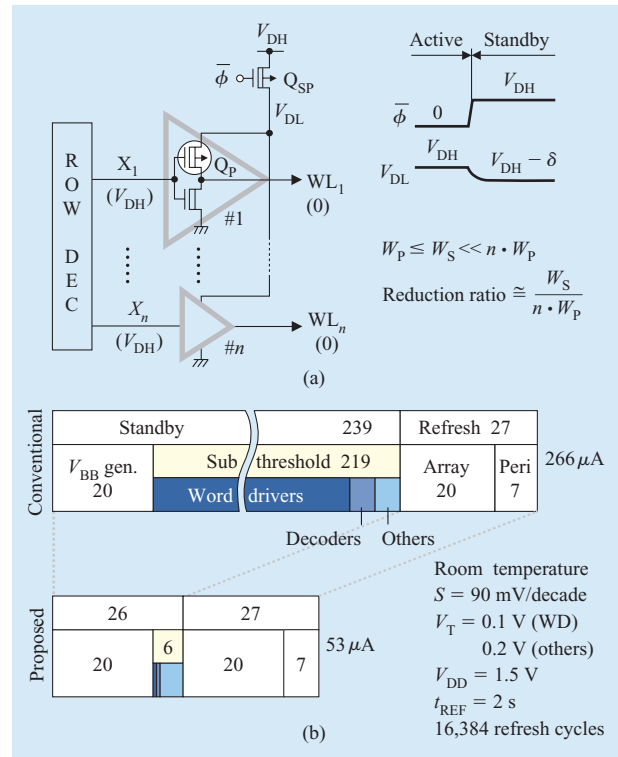


Figure 22

Gate-source voltage ( $V_{GS}$ ) self-reverse biasing applied to a 256-Mb DRAM: (a) Application to word drivers; (b) reduction of retention current.  $W_S$  and  $W_P$  denote the respective channel widths of  $Q_{SP}$  and  $Q_P$ .  $V_T$  is defined by current density of 10 nA/15  $\mu$ m. Reproduced from [70] with permission; © 1993 IEEE.

reduction scheme (e.g.,  $V_{GS}$  self-reverse biasing) in advance. As for input nodes, which are not predictable, the level-fixing input buffer (Figure 20) [15] can force the internal node voltages to be predictable. In standby mode (signal STANDBY is at high level), internal nodes including  $a_i$ ,  $\bar{a}_i$ , and the following-stage outputs are forced to be at predetermined levels, irrespective of input node  $A_i$ . Similar techniques are applied to logic LSIs, though their node voltages are usually unpredictable because they contain registers or latches to retain internal states. Latches (Figure 21) [59] that fix the output level while retaining the latched data are effective in reducing  $I_{leak}$  in sleep mode. Level-fixing flip-flops [67] combined with self-reverse biasing [15], power switches [60], and level holders [18] enable quick recovery from sleep mode. These techniques, in turn, can be applied to RAM peripheral circuits with registers or latches.

#### Slow cycle

RAMs feature a slow cycle  $t_{RC}$  compared with random logic gates, and this allows each circuit to be active for only a



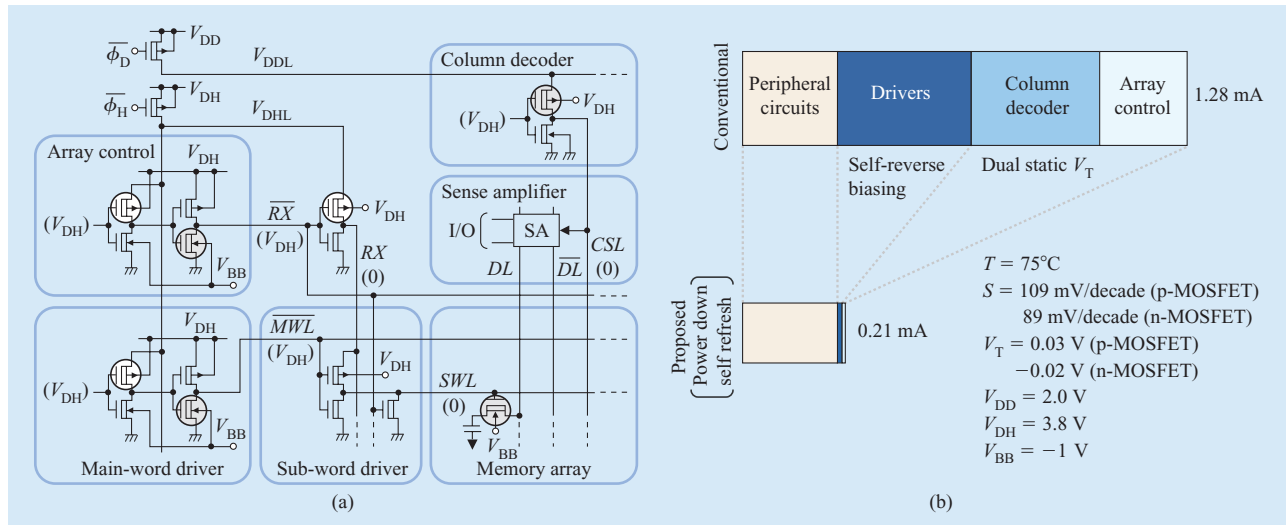


Figure 23

Various leakage-reduction schemes applied to 256-Mb SDRAM [57]: (a) Application to array-associated circuitry; (b) leakage-current reduction.  $V_T$  is defined by a current density of 10 nA/15  $\mu\text{m}$ . The peripheral circuits component is from peripheral MOSFETs without substrate bias.

short period within the “long” memory cycle, leaving additional time to control the subthreshold current. This is true for DRAM row circuits, which are slow enough to accept leakage controls. However, the column circuits in modern DRAMs (Figure 19) feature a fast burst cycle and unpredictable circuit operation (every column may be selected during the memory cycle). Therefore, it is difficult to reduce  $I_{\text{leak}}$  in column circuits in the active mode. This is the case for high-speed SRAMs and logic LSIs.

*Use of robust circuits*

RAMs do not use leakage-sensitive circuits, such as dynamic NOR gates, that require a level keeper to prevent malfunctions caused by leakage [68]. The decoders of modern CMOS DRAMs consist of dynamic (for the row) and static (for the column) NAND gates to reduce the power (Figure 19). NAND decoders discharge only one output node in a selected decoder, while the NOR decoders used in the n-MOS era discharged all output nodes in decoders, except for the selected one.

In contrast, it is difficult to reduce  $I_{\text{leak}}$  in random logic circuits because of the noniterative circuit topology, higher activation ratio, unpredictable node states, and faster cycle. Dual static  $V_T$  [53], the stack effect in NAND gates described above, and circuit reordering [69] are effective to some extent in reducing  $I_{\text{leak}}$  in the standby mode of logic LSIs. However, reducing  $I_{\text{leak}}$  in random logic circuits in the active mode is more difficult. The only scheme that has been reported thus far is dual static  $V_T$ , though it has

limited reduction efficiency because of the limited  $V_T$  difference, as previously explained. More effective schemes have yet to be discovered.

*Applications to DRAM standby mode*

The reduction of subthreshold leakage current applied to iterative circuit blocks, such as a word-driver block, is extremely important in memory design. For example, a low- $V_T$  p-MOS switch [ $Q_{\text{sp}}$  in Figure 22(a)] [14, 15] shared with the  $n$  word drivers of a 256-Mb DRAM [70] enables the common power line,  $V_{\text{DL}}$ , to drop by  $\delta$  as a result of the total subthreshold current flow of  $nI$  when the switch is off in standby mode. As it provides each p-MOS driver,  $Q$ , with a  $\delta$  self-reverse bias, the subthreshold current,  $I$ , eventually decreases. Hence, even if an on-chip charge pump for the raised supply  $V_{\text{DH}}$  necessary for DRAM word-line bootstrapping suffers from poor output-current drivability, the  $V_{\text{DH}}$  is well regulated. In the active mode, the selected word line is driven after  $V_{\text{DL}}$  is connected to a supply voltage,  $V_{\text{DH}}$ , by turning on  $Q_{\text{sp}}$ . Here, the channel width of  $Q_{\text{sp}}$  can be reduced to an extent comparable to that of  $Q$  without a speed penalty because of the low activation ratio,  $1/n$ , of the drivers. In a 256-Mb chip, a  $\delta$  as small as 0.25 V reduced the standby subthreshold current of word drivers and decoders by two decades [Figure 22(b)] without inflicting penalties in terms of speed and area.

Another example is shown in Figure 23(a). This 256-Mb SDRAM [57] with a hierarchical word-line structure

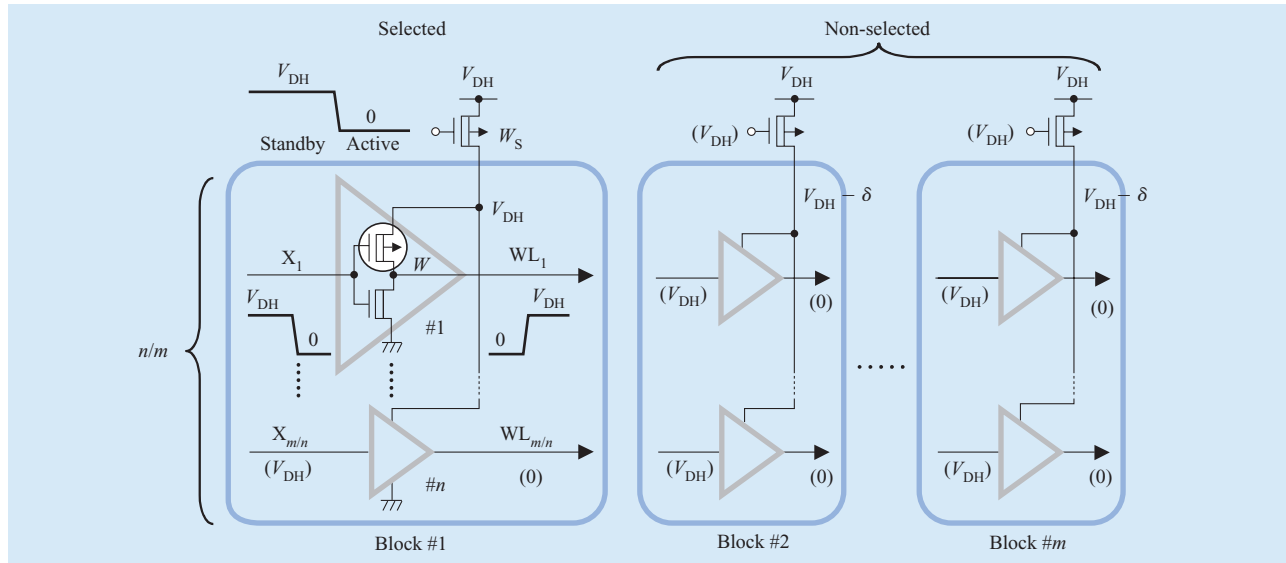


Figure 24

Active leakage current reduction with partial activation of multi-divided subarray [18].

utilizes the self-reverse biasing described above combined with “pseudo” multiple static  $V_T$  using substrate biasing. The circled MOSFETs in the figure are in the subthreshold region during standby mode. Here, self-reverse biasing is applied only to p-MOSFETs (open circles) that produce larger subthreshold current. This is because p-MOSFETs have larger total channel width and larger subthreshold swing due to the buried-channel MOSFET structure. The n-MOSFETs (shaded circles) and the p-MOSFETs in the column decoder have higher  $V_T$  due to the respective well bias  $V_{BB}$  and  $V_{DH}$ . By combining both schemes, the total subthreshold leakage current in the power-down/self-refresh mode is reduced to one sixth, as **Figure 23(b)** shows. The current can be further reduced by applying both schemes to the peripheral circuits.

#### Applications to DRAM active mode

In the future, with a further reduction in  $V_T$ , the subthreshold leakage current,  $I_{DC}$ , will exceed the capacitive current,  $I_{AC}$ , and eventually dominate the total active current,  $I_{ACT}$ , of the chip [Figure 6(b)], as pointed out as early as 1993 [18, 71].  $V_{GS}$  back-biasing applied to an iterative circuit block, which is divided into  $m$  sub-blocks, each consisting of  $n/m$  circuits (**Figure 24**), confines the currents to that of a single selected sub-block [18]. This is because all nonselected sub-blocks have no substantial subthreshold current due to  $V_{GS}$  back-biasing (Figure 22) when the switch of the selected sub-block, including the selected word line, is turned on while the

others remain off. The above-mentioned multi-static  $V_T$  also reduces current. The subthreshold currents of low- $V_T$  circuits on the critical path are reduced by combining power switches and high- $V_T$  level holders (**Figure 25**) [18, 72]. The power switch goes off just after evaluating the input of the low- $V_T$  circuit and holding the evaluated output at the holder. This prevents the output from discharging, allowing the switch to quickly turn on at the necessary time to prepare for the next evaluation. This is a good example of the principle of avoiding large voltage swings with heavily capacitive loads. In fact, it has been reported that these circuits could reduce the active current of a hypothetical 16-Gb DRAM [18, 71] from 1.2 A to 0.1 A (**Figure 26**), although their effectiveness with an actual chip has not yet been verified.

#### 5. Speed variations and other issues with peripheral circuits

Other key peripheral circuits are sense amplifiers and low-voltage supporting circuits, such as level shifters, stress-release I/O circuits, and on-chip supply-voltage generators in RAM chips (Figure 3). They play important roles in the stability and speed of RAMs. However, well-known logic-gate blocks in peripheral circuits are also important in terms of suppression of speed variations, as explained earlier. Power management is essential for high-speed, low-power designs of the blocks. Testing methodology that is relevant to leakage currents is also a major area of concern.

### Sense amplifiers

Sense amplifiers (SAs) are always slow because they manage a small signal, thus requiring high-speed design achieved by reducing speed variations. The design of SAs [4], which usually have a cross-coupled circuit configuration in terms of low power and small area, can be different for DRAMs and SRAMs. This is because the necessary size, the number in a chip, and the circuit operation are usually different. DRAMs feature a huge number of tiny SAs in a chip, because one SA must be placed at each data line due to refresh requirements. In addition, in the standard mid-point (half- $V_{DD}$ ) sensing of DRAMs [4], the SA must operate at the lowest voltage (i.e., half- $V_{DD}$ ) in the chip, despite the resulting halved data-line power without a dummy cell and with a low-noise array [4]. As a result, the statistically large  $V_T$  variations,  $\sigma(V_T)$ , and low-voltage operation slow down sensing with a wide spread in speed. Increasing the size of SA MOSFETs to reduce  $\sigma(V_T)$  and using redundancy and/or ECC to prevent SAs from acquiring an excessively large  $\delta V_T$  are effective solutions that are similar to those associated with the  $V_T$ -mismatch issue previously explained in the subsection on cell signal charge in Section 2. In overdrive sensing [73, 74], this problem is solved by applying a higher voltage solely to SA inputs by isolating the data line from the SA or by capacitive coupling. Using additional capacitors may be acceptable in e-DRAMs, where area is of less concern. The recently presented full- $V_{DD}$  (or ground) sensing with a dummy cell [5], which is a revival of the kind of sensing done during the n-MOS DRAM era of the 1970s, solves the problem with a raised voltage (i.e.,  $V_{DD}$ ).

SRAMs have a small number of SAs on a chip, although they must be highly sensitive for a higher speed. Thus, in addition to some of the above solutions for DRAMs, a low-voltage current SA [75] may be acceptable despite the increase in area.

### Low-voltage supporting circuits

High-speed level shifters that are proposed for SoCs [76, 77] and bridge the internal low-voltage core and high-voltage I/O circuits could be used for RAMs. Low-cost stress-release I/O circuits [78–80] that manage the high voltage at the interface with a single thin  $t_{OX}$  are also important. On-chip supply-voltage generators [4] continue to be essential in the stable operation of RAM cells with high supply voltages and in standardizing the power supply of standalone RAMs. In addition, they reduce subthreshold currents with multi- $V_T$  (Figure 23) and speed variations at lower external supply voltages, as discussed below. Key issues are a high efficiency of voltage conversion, a high degree of accuracy in the output

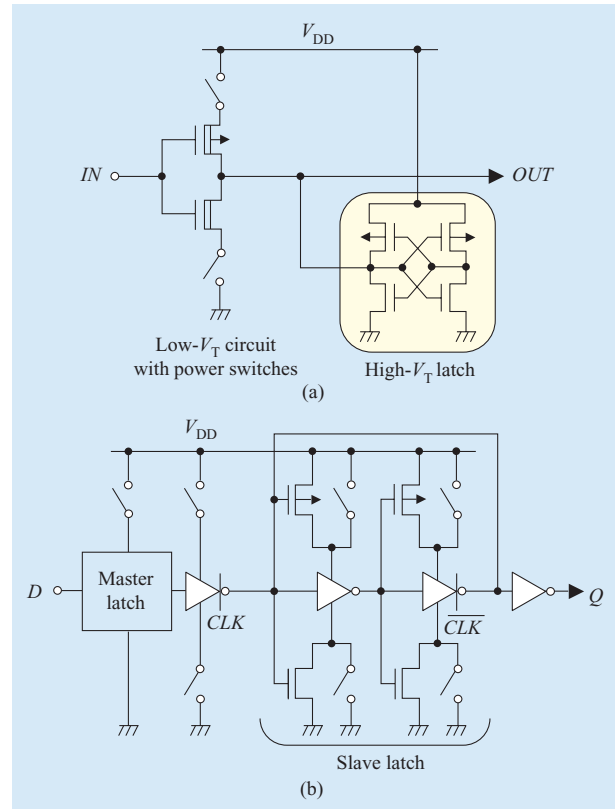


Figure 25

Subthreshold leakage current reduction of input-unpredictable circuit: (a) Power switch with latch. Reproduced from [18] with permission; ©1994 IEEE. (b) Its application to flip-flop [72]. Inverters and clocked inverters consist of low- $V_T$  MOSFETs.

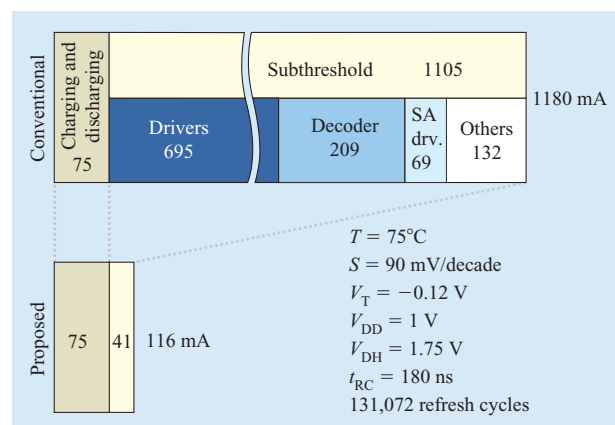


Figure 26

Active current reduction in hypothetical 16-Gb DRAM.  $V_T$  is defined by a current density of  $10 \text{ nA}/5 \mu\text{m}$ . Reproduced from [18] with permission; © 1994 IEEE.

voltage, low power during the standby period, and a low cost of implementation [27].

### **Power management**

Power management is a solution to suppress speed variations and further reduce the power dissipation of power-aware systems through static and dynamic control of supply voltages. Power management can also effectively reduce subthreshold currents with  $V_{BB}$  control, as mentioned earlier. Many schemes have thus far been proposed. The following subsections give a brief discussion of power-management problems that DRAM designers have experienced, followed by various viewpoints on power-management schemes that have been proposed by logic designers principally for SoCs.

In the past, DRAM designers encountered numerous problems that occurred even in static or quasi-static  $V_{BB}$  and  $V_{DD}$ . It is well known that the DRAM has been the only large-volume production LSI using a substrate bias that is supplied from an on-chip  $V_{BB}$  generator. In the n-MOS DRAM era, when a quasi-static  $V_{BB}$  was supplied to the p-type substrate of the whole chip (i.e., both array and periphery), the generator caused instabilities (surge current [55] or a degraded  $V_{BB}$  level [4]) at power-on and during burn-in high-voltage stress tests, and shortened the refresh time of cells due to minority-carrier injection to cells [4]. Poor current drivability of the generator consisting of charge pumps, a large substrate current generated from the peripheral circuits, and the substrate structure were mainly responsible for the instabilities. Even so, DRAM designers were fortunate because both the static bias setting of a deep  $V_{BB}$  of about  $-2$  V to  $-3$  V and a sufficiently high  $V_T$  of about  $0.5$  V allowed stable chip operation with small changes in  $V_T$ , even with quite large quasi-static  $V_{BB}$  variations and  $V_{BB}$  noise [4]. In the CMOS era, substrate bias was removed from peripheral circuits primarily to eliminate instabilities caused by the generator and has only been supplied to the array to ensure stable operation.

Even a bump as small as  $\pm 10\%$   $V_{DD}$  made dynamic circuits unstable during the n-MOS era. This was due to a charge being trapped at floating nodes when voltage bumps were applied, causing malfunctions at the next cycle. Note that almost all peripheral circuits and DRAM cells were dynamic. Thus, a small diode-connected n-MOS (i.e., level keeper) was connected to the floating nodes of peripheral circuits to allow trapped charges to escape. However, bumps degraded the voltage margin of n-MOS cells, calling for grounded-plate cell capacitors [4] as a partial solution. Even in the CMOS era, memory cells, sensing relevant circuits (such as data-line precharge circuits and sense amplifiers) and row decoders/drivers were still dynamic, while other peripheral circuits have

been static. Half- $V_{DD}$  sensing [81] (coupled with a half- $V_{DD}$  cell plate and a boosted word line) has been a circuitry solution because the margins of DRAM cells and the relevant sensing circuits are maintained wide despite voltage bumps. A CMOS feedback level keeper that is familiar to logic designers has been widely used for other dynamic circuits.

### **Static control of power-supply voltages**

Static control is effective in suppressing speed variations of logic circuits while preserving stability of memory cells and memory-cell-relevant circuits. When  $V_{BB}$  or internal  $V_{DD}$  is statically controlled on the basis of parameter variations, inter-die speed variations can be suppressed, although intra-die speed variations remain unimproved. Negative effects, if any, when supply voltages are controlled statically or quasi-statically could be managed, as memory designers have done thus far. Controlling  $V_{BB}$  with an on-chip  $V_{BB}$  generator to adjust  $V_T$  (the basic idea dates back to 1976 [62, 82, 83]) could be widely used to suppress the variations if the previously discussed drawbacks are rectified. Controlling forward  $V_{BB}$ , however, is more effective in reducing speed variations [84–86] because the  $V_T$ - $V_{BB}$  characteristics are more sensitive to  $V_{BB}$  [4]. For example, controlling forward  $V_{BB}$  reduced  $V_T$  variations in logic circuits and improved speed of operations by 10% [85]. If a forward  $V_{BB}$  is used, however, the requirements to suppress noise become more stringent, calling for a uniform distribution of the forward  $V_{BB}$  throughout the chip [27]. Additional current consumption, in the form of bipolar current induced by the forward  $V_{BB}$ , is another matter [85] that must be considered.

Control of internal  $V_{DD}$  with an on-chip voltage-down converter (i.e., series regulator) [4] seems to be more practical, because the instabilities discussed above are not involved. In fact, a  $V_{DD}$  control with both an off-chip buck converter and an internal-delay-detecting circuit [87] reduced the variation between speeds of the worst and best design conditions from five times to  $\pm 20\%$  at  $0.5$  V. However, the use of an on-chip voltage-down converter instead of the buck converter may be more practical because designs of the converter are simpler and have been well established in DRAM designs despite a lower conversion efficiency.

### **Dynamic control of power-supply voltages**

Dynamic control reduces power dissipation and subthreshold currents. However, the problems described above might be compounded and become serious if dynamic control of  $V_{DD}$  and/or  $V_{BB}$  were applied to RAM chips, because they involve wide and dynamic changes in supply voltages and extremely low  $V_T$ . Nevertheless, many attempts have been made, although only for the logic

blocks of SoCs. Unfortunately, RAM cells and their relevant circuits are incompatible with dynamic controls, and thus they should at least be “quiet.” Moreover, they must operate at a higher  $V_{DD}$ . Their inherently small voltage margins are responsible for the requirements for the quiet and higher- $V_{DD}$  operation, as previously explained. Thus, as long as the controls never cause detrimental effects to RAM cells and their relevant circuits, some of them could be applied to parts of peripheral logic circuits (e.g., static circuits) in RAM chips or RAM blocks in SoCs. Note that SRAM blocks using full CMOS SRAM cells may accept dynamic voltage controls to some extent because of wide voltage margins, although care should be taken if dynamic sensing schemes are adopted.

Power switches [88] completely cut leakage currents of internal core circuits, although they incur a long recovery time on heavily capacitive internal power lines, as was explained in the subsection on circuit applications in Section 4. Dynamic voltage scaling (DVS) [89, 90], in which the clock frequency and  $V_{DD}$  vary dynamically in response to the computational load, provides reduced energy consumption per process during periods when few computations are performed, while still providing peak performance when required. Note that the highest  $V_{DD}$  and lowest  $V_{DD}$  that DVS can accept must be determined by the breakdown voltage of MOSFETs and the stability of RAM cells, respectively. This approach, however, becomes less effective in the low- $V_{DD}$  era because the range across which it is possible to vary  $V_{DD}$  becomes narrower. In addition, successful operation over a wide range of  $V_{DD}$  requires the accurate tracking of all circuit delays. Furthermore, applying DVS would make dynamic circuits (e.g., e-DRAMs) unstable without a level keeper [90], although resultant instabilities depend on the changing rate of  $V_{DD}$  and clock frequency.

For partially depleted (PD) SOIs, a wide changing of  $V_{DD}$  may cause additional instabilities due to the floating-body effect. DVS does not reduce subthreshold currents; these currents are reduced by elastic- $V_T$  CMOS [52], where the clock frequency,  $V_{DD}$ , and  $V_{BB}$  are all dynamically varied. However, substrate noise may be coupled from the  $V_{DD}$  power line when  $V_{DD}$  is changed, which is hazardous in an on-chip  $V_{BB}$  scheme. The cost and complexity of design are additional problems.

System-level low-power techniques introduced into a SoC would be effective if the problems described above could be solved. For example, ChipOS [91] was introduced to specify the acceptable maximum power and thus, maximum junction temperature. The power of the logic block for each sub-block is managed by controlling the gated clock and power switch to achieve a given power budget. In autonomous decentralized low-power systems [92, 93], the frequency, supply voltage, substrate bias

voltage, and power switch of each sub-block are all controlled by the system, according to its supplied processing load, to achieve the minimum power consumption. Even in this scheme, high-speed controls (e.g., for fast wake-up) of subthreshold currents of selected and nonselected sub-blocks would be essential.

### Testing

Testing of low-voltage RAMs is problematic. A large subthreshold current makes it difficult to discriminate between defective and non-defective  $V_{DD}$  currents (i.e.,  $I_{DDQ}$  currents), and thereby poses a problem in the  $I_{DDQ}$  testing of low-voltage CMOS circuits.  $I_{DDQ}$  testing with the application of a reverse  $V_{BB}$  [94] is effective when low-temperature measurement and multi- $V_T$  design are combined. Lowering  $V_{DD}$  only at detection is also important because it dramatically reduces GIDL currents. The unusual temperature dependence of speed (even nullified) at a lower  $V_{DD}$  [95, 96] is another concern in speed testing.

## 6. Future prospects

On the basis of the above, we present future perspectives on low-voltage RAMs in terms of devices and processes, memory cells, peripheral circuits, and architectures.

### Devices and processes

#### Device structure of RAM chips

In the near future, RAMs must unavoidably take at least a dual- $t_{OX}$ , dual- $V_T$ , and dual- $V_{DD}$  approach because of different requirements between RAM cells and peripheral circuits, as discussed in the subsections on cell signal charge and leakage currents in Section 2. RAM cells require an ever-higher  $V_T$  (Figure 5) and thus, a high  $V_{DD}$  and thick  $t_{OX}$  for stable and reliable operation. In contrast, peripheral circuits (or logic blocks on a SoC) require a low  $V_{DD}$ , low  $V_T$ , and thus, thin  $t_{OX}$  for fast and low-power operations, according to ITRS trends [8]. For a higher I/O interface voltage, a triple  $t_{OX}$  would be popular.

#### Low-leakage currents

Even one of the most-advanced schemes (Figure 10) would be less effective for lower-voltage, larger-capacity SRAMs. The resultant total current is as large as 1.6  $\mu\text{A}$  even for a memory capacity as small as 16 Mb—even if a large  $V_T$ , a thick  $t_{OX}$ , and an offset source driving are all combined. Thus, much larger  $V_T$  and thicker  $t_{OX}$  are needed in the future, calling for new devices such as fully depleted (FD) SOIs with a reduced S-factor and new gate-insulator materials. In addition, lowering  $V_{DD}$  while keeping the voltage swing the same to preserve the effectiveness of the scheme increases SER to unacceptable levels because of decreased  $Q_s$  in the standby mode,

calling for soft-error-immune devices as well as on-chip ECC circuits.

#### **Low voltage and high speed**

PD-SOIs [97] have been successfully used for products such as MPUs because they improve the performance of standard digital logic by 20–35% over the comparable bulk process due to reduced capacitance. Major concerns with PD-SOIs, however, are the instabilities [4, 97] caused by the floating body. In particular, the resulting  $V_T$  variations degrade margins of cells and their relevant circuits, and the degradation is further enhanced at a lower  $V_{DD}$ . For SRAMs, some solutions have been proposed. These include reducing the number of cells connected to one column [98] to lower the accumulated subthreshold leakage from nonselected cells. A body contact applied to the paired MOSFETs of a sense amplifier [99] reduces sense-amplifier offset. A body-tied substrate with partial trench isolation [100, 101] is a solution to significantly improve immunity against soft errors while eliminating instabilities. The floating body in DRAMs degrades data retention time in the 1-T DRAM cell [102]. A combination of bulk for the DRAM cell array and an SOI for the peripheral circuits [103] is a solution despite the costly substrate structure.

The use of a dynamic threshold MOSFET (DTMOS) [104], which is built with the body connected to the gate and thus enables a non-floating body, is attractive in terms of low-voltage operation and the suppression of speed variations. This lowers the upper limit of the  $V_{DD}$  to less than 0.5 V, even at room temperature, because of the rapid increase in pn-forward current [85]. However, the feature of self-corrective  $V_T$  [85, 87] that DTMOS provides can suppress speed variations.

Although the concept of DTMOS was originally proposed with PD-SOIs despite the highly resistive body, it has also been realized with bulk MOSFETs with a low-resistive body [87]. Coupled with an internal  $V_{DD}$  control, the DTMOS with bulk MOSFETs reduced the delay variation (i.e., delay difference between the worst and best design conditions) to one-fiftieth at 0.5 V. In addition, it realized a drive current three times greater than a conventional CMOS, while reducing the subthreshold current to two orders of magnitude.

FD-SOIs are also attractive in low-voltage operation because of the reduced  $S$ -factor, a small junction capacitance, small body-bias effects, and a small layout area. Thus, excellent performances [87, 105–107] have been achieved with multi- $V_T$  (dual/triple) FD-SOI, despite low voltages (0.3–0.5 V) and still large (0.25- $\mu\text{m}$ ) FD-SOI processes. In the 0.1- $\mu\text{m}$  or less era, however, we need to reduce additional  $V_T$  variations [108], if any, caused by thickness variations of the thin body and to attain multi- $V_T$  in specific MOSFETs to reduce subthreshold currents,

although uses of special gate materials [109] and gate doping [110] have been proposed. Note that realizing multi- $V_T$  through dynamic  $V_{BB}$  is impossible with FD-SOIs because of the lack of a body.

Because it seems unlikely that device and process solutions will be developed in time, the pace at which  $V_{DD}$  is being lowered should be slowed so that larger MOSFETs are acceptable. Hence, vertical MOSFETs [111] that accept large channel length and  $t_{OX}$  without sacrificing density might be effective. Vertical MOSFETs may also reduce RAM cell areas [41]. If the above attempts are unsuccessful, low-temperature bulk CMOS [112] may have a resurgence in the future.

#### **Memory cells**

In addition to small, high-speed ECC circuits, new RAM cells such as gain cells are indispensable, as explained in the subsection on DRAM cells in Section 3. In the long run, however, high-speed, high-density nonvolatile RAMs show strong potential for use as low-voltage memories. In particular, leakage-free and soft-error-free structures and the nondestructive read-out and non-charge-based operations that they could provide are attractive in terms of achieving fast cycle times, low power with zero standby power, and stable operation, even at the lower  $V_{DD}$ . Simple planar structures, if possible, would cut costs. In this sense, magnetic RAMs (MRAMs) [113] and Ovonic Unified Memories\*\* (OUMs\*\*) [114] are appealing propositions. In MRAMs, one major drawback remains, which is to reduce the magnetic field needed to switch the magnetization of the storage element, while in OUMs, managing the proximity heating of the cell is an issue. In addition, the scalabilities and stability required to ensure nonvolatility still remain unresolved because development is still in its early stages.

#### **Peripheral circuits and architectures**

As far as RAMs are concerned, the subthreshold currents in the active mode could be reduced by improving the above-described CMOS circuits, unless they are too fast. In high-speed RAMs, such as fast SRAMs or high-speed column-mode DRAMs, however, current reduction is extremely difficult, as discussed in the subsection on applications to RAMs in Section 4. This suggests that a high-speed SoC will suffer from incredibly high power dissipated by its random logic gates because it may remain impossible to control subthreshold currents from these logic gates at a sufficiently high speed. Hence, the number of gates must be reduced. This implies that new SoC architectures will be required, such as memory-rich SoCs, which effectively reduce the subthreshold current. In addition to new architectures, low-power techniques learned from “old circuits,” such as bipolar, BiCMOS, E/D

MOS, capacitive boosting, CML circuits, and even I<sup>2</sup>Ls, might be necessary.

## 7. Summary

This paper reviewed technology trends in low-voltage DRAMs and SRAMs and clarified the challenges facing low-voltage RAMs in terms of cell signal charge, necessary threshold voltage,  $V_T$ , and  $V_T$  variations in the MOSFETs of RAM cells and sense amplifiers, and leakage current. It then discussed developments in conventional RAM cells and emerging cells, such as DRAM gain cells and leakage-immune SRAM cells, from the viewpoints of cell area, operating voltage, and the subthreshold and gate-tunnel currents of MOSFETs. The concepts behind reducing subthreshold currents that have been proposed to date and the features of RAMs with respect to reducing subthreshold currents were then summarized. After that, their applications to RAM circuits to reduce subthreshold currents in standby and active modes were discussed, exemplified by DRAMs. The paper then discussed design issues in other peripheral circuits, such as sense amplifiers, I/O circuits, and on-chip power-supply generators, and it investigated the suppression of speed variations and power reductions through power management and testing. With respect to the above, future prospects were considered, with an emphasis on needs for high-speed nonvolatile RAMs, subthreshold-current reduction for high-speed active mode, and memory-rich SoC architectures.

## Acknowledgment

The authors would like to thank Dr. D. Hisamoto and Dr. R. Tsuchiya for their valuable discussions on SOI and bulk CMOS characteristics.

\*\*Trademark or registered trademark of Monolithic System Technology Inc. or Ovonyx, Inc.

## References

1. H. Yoon, J. Y. Sim, H. S. Lee, K. N. Lim, J. Y. Lee, N. J. Kim, K. Y. Kim, S. M. Byun, W. S. Yang, C. H. Choi, H. S. Jeong, J. H. Yoo, D. I. Seo, K. Kim, B. I. Ryu, and C. G. Hwang, "A 4Gb DDR SDRAM with Gain-Controlled Pre-Sensing and Reference Bitline Calibration Schemes in the Twisted Open Bitline Architecture," *ISSCC Digest of Technical Papers*, February 2001, pp. 378–379.
2. D. H. Kim, S. J. Kim, B. J. Hwang, S. H. Seo, J. H. Choi, H. S. Lee, W. S. Yang, M. S. Kim, K. H. Kwak, J. Y. Lee, J. Y. Joo, J. H. Kim, K. Koh, S. H. Park, and J. I. Hong, "Highly Manufacturable 32Mb ULP-SRAM Technology by Using Dual Gate Process for 1.5V Vcc Operation," *Symposium on VLSI Technology, Digest of Technical Papers*, June 2002, pp. 118–119.
3. U.-R. Cho, T.-H. Kim, Y.-J. Yoon, J.-C. Lee, D.-G. Bae, N.-S. Kim, K.-Y. Kim, Y.-J. Son, J.-S. Yang, K.-I. Sohn, S.-T. Kim, I.-Y. Lee, K.-J. Lee, T.-G. Kang, S.-C. Kim, K.-S. Ahn, and H.-G. Byun, "A 1.2V 1.5Gb/s 72Mb DDR3 SRAM," *ISSCC Digest of Technical Papers*, February 2003, pp. 300–301.
4. K. Itoh, *VLSI Memory Chip Design*, Springer-Verlag, New York, 2001.
5. J. Barth, D. Anand, J. Dreibelbis, and E. Nelson, "A 300MHz Multi-Banked eDRAM Macro Featuring GND Sense, Bit-Line Twisting and Direct Reference Cell Write," *ISSCC Digest of Technical Papers*, February 2002, pp. 156–157.
6. D. Weiss, J. J. Wu, and V. Chin, "The On-Chip 3MB Subarray Based 3rd Level Cache on an Itanium Microprocessor," *ISSCC Digest of Technical Papers*, February 2002, pp. 112–113.
7. K. Itoh, T. Watanabe, S. Kimura, and T. Sakata, "Reviews and Prospects of High-Density RAM Technology," *Proceedings of CAS 2000*, October 2000, Sinaia (Romania), pp. 13–22.
8. *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2001 Edition.
9. K. Itoh, "Reviews and Prospects of Deep Sub-Micron DRAM Technology," *Extended Abstracts of the International Conference on Solid-State Devices and Materials*, August 1991, pp. 468–471.
10. M. Aoki, J. Etoh, K. Itoh, S. Kimura, and Y. Kawamoto, "A 1.5V DRAM for Battery-Based Applications," *ISSCC Digest of Technical Papers*, February 1989, pp. 238–239.
11. Y. Nakagome, Y. Kawamoto, H. Tanaka, K. Takeuchi, E. Kume, Y. Watanabe, T. Kaga, F. Murai, R. Izawa, D. Hisamoto, T. Kisu, T. Nishida, E. Takeda, and K. Itoh, "A 1.5-V Circuit Technology for 64Mb DRAMs," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1990, pp. 17–18.
12. J. Etoh, K. Itoh, Y. Kawajiri, Y. Nakagome, E. Kume, and H. Tanaka, "Large Scale Integrated Circuit for Low Voltage Operation," U.S. Patent 5,297,097, March 1994.
13. Y. Nakagome, K. Itoh, M. Isoda, K. Takeuchi, and M. Aoki, "Sub-1-V Swing Bus Architecture for Future Low-Power ULSIs," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1992, pp. 82–83.
14. T. Kawahara, M. Horiguchi, Y. Kawajiri, G. Kitsukawa, T. Kure, and M. Aoki, "Subthreshold Current Reduction for Decoded-Driver by Self-Reverse Biasing," *IEEE J. Solid-State Circuits* **28**, No. 11, 1136–1144 (November 1993).
15. M. Horiguchi, T. Sakata, and K. Itoh, "Switched-Source-Impedance CMOS Circuit for Low Standby Subthreshold Current Giga-Scale LSI's," *IEEE J. Solid-State Circuits* **28**, No. 11, 1131–1135 (November 1993).
16. D. Takashima, S. Watanabe, H. Nakano, Y. Oowaki, K. Ohuchi, and H. Tango, "Standby/Active Mode Logic for Sub-1-V Operating ULSI Memory," *IEEE J. Solid-State Circuits* **29**, No. 4, 441–447 (April 1994).
17. 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* **30**, No. 8, 847–854 (August 1995).
18. T. Sakata, K. Itoh, M. Horiguchi, and M. Aoki, "Subthreshold-Current Reduction Circuits for Multi-Gigabit DRAM's," *IEEE J. Solid-State Circuits* **29**, No. 7, 761–769 (July 1994).
19. K. Itoh, K. Sasaki, and Y. Nakagome, "Trends in Low-Power RAM Circuit Technologies," *Proc. IEEE* **83**, No. 4, 524–543 (April 1995).
20. E. Ibe, "Current and Future Trend on Cosmic-Ray-Neutron Induced Single Event Upset at the Ground Down to 0.1-Micron-Devices," presented at the Svedberg Laboratory Workshop on Applied Physics, May 2001, Uppsala, Sweden.

21. Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," *Proc. IEEE* **85**, No. 4, 486–504 (April 1997).
22. S. Hong, S. Kim, J.-K. Wee, and S. Lee, "Low-Voltage DRAM Sensing Scheme with Offset-Cancellation Sense Amplifier," *IEEE J. Solid-State Circuits* **37**, No. 10, 1356–1360 (October 2002).
23. J. Y. Sim, K. W. Kwon, J. H. Choi, S. H. Lee, D. M. Kim, H. R. Hwang, K. C. Chun, Y. H. Seo, H. S. Hwang, D. I. Seo, and S. I. Cho, "A 1.0V 256Mb SDRAM with Offset-Compensated Direct Sensing and Charge-Recycled Precharge Scheme," *ISSCC Digest of Technical Papers*, February 2003, pp. 310–311.
24. H. L. Kalter, C. H. Stapper, J. E. Barth Jr., J. DiLorenzo, C. E. Drake, J. A. Fifield, G. A. Kelley, Jr., S. C. Lewis, W. B. van der Hoeven, and J. A. Yankosky, "A 50-ns 16-Mb DRAM with a 10-ns Data Rate and On-Chip ECC," *IEEE J. Solid-State Circuits* **25**, No. 5, 1118–1128 (October 1990).
25. K. Osada, Y. Saitoh, E. Ibe, and K. Ishibashi, "16.7fA/Cell Tunnel-Leakage-Suppressed 16-Mbit SRAM Based on Electric-Field-Relaxed Scheme and Alternate ECC for Handling Cosmic-Ray-Induced Multi-Errors," *ISSCC Digest of Technical Papers*, February 2003, pp. 302–303.
26. K. Itoh, "Reviews and Prospects of Low-Power Memory Circuits" (invited), *Low-Power CMOS Design*, A. Chandrakasan and R. Brodersen, Eds., Wiley-IEEE Press, Hoboken, NJ, 1998, pp. 313–317.
27. K. Itoh and H. Mizuno, "Low-Voltage Embedded-RAM Technology: Present and Future," *Proceedings of the 11th IFIP International Conference on Very Large Scale Integration of Systems-on-Chip (VLSI-SOC'01)*, December 2001, pp. 277–288.
28. O. Takahashi, S. Dhong, M. Ohkubo, S. Onishi, R. Dennard, R. Hannon, S. Crowder, S. Iyer, M. Wordeman, B. Davari, W. B. Weinberger, and N. Aoki, "1GHz Fully Pipelined 3.7ns Address Access Time 8k × 1024 Embedded DRAM Macro," *ISSCC Digest of Technical Papers*, February 2000, pp. 396–397.
29. *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2002 Edition.
30. D. J. Frank, "Power-Constrained CMOS Scaling Limits," *IBM J. Res. & Dev.* **46**, No. 2/3, 235–244 (March 2002).
31. T. Inukai and T. Hiramoto, "Suppression of Stand-By Tunnel Current in Ultra-Thin Gate Oxide MOSFETs by Dual Oxide Thickness MTCMOS," *Extended Abstracts of the International Conference on Solid-State Devices and Materials*, August 1999, pp. 264–265.
32. S. Wei Sun and P. G. Y. Tsui, "Limitation of CMOS Supply-Voltage Scaling by MOSFET Threshold-Voltage Variation," *Proceedings of the CICC*, May 1994, pp. 267–270.
33. C. J. Radens, U. Gruening, J. A. Mandelman, M. Seitz, T. Dyer, D. Lea, D. Casarotto, L. Clevenger, L. Nesbit, R. Malik, S. Halle, S. Kudelka, H. Tews, R. Divakaruni, J. Sim, A. Strong, D. Tibbel, N. Arnold, S. Bukofsky, J. Preuninger, G. Kunkel, and G. Bronner, "A 0.135 $\mu\text{m}^2$  6F<sup>2</sup> Trench-Sidewall Vertical Device Cell for 4 Gb/16 Gb DRAM," *Symposium on VLSI Technology, Digest of Technical Papers*, June 2000, pp. 80–81.
34. F. Hofmann and W. Rosner, "Surrounding Gate Select Transistor for 4F<sup>2</sup> Stacked Gbit DRAM," *Proceedings of ESSDERC*, September 2001, pp. 131–134.
35. T. Takahashi, T. Sekiguchi, R. Takemura, S. Narui, H. Fujisawa, S. Miyatake, M. Morino, K. Arai, S. Yamada, S. Shukuri, M. Nakamura, Y. Tadaki, K. Kajigaya, K. Kimura, and K. Itoh, "A Multi-Gigabit DRAM Technology With 6F<sup>2</sup> Open-Bit-Line Cell Distributed Over-Driven Sensing and Stacked-Flash Fuse," *ISSCC Digest of Technical Papers*, February 2001, pp. 380–381.
36. T. Sekiguchi, K. Itoh, T. Takahashi, M. Sugaya, H. Fujisawa, M. Nakamura, K. Kajigaya, and K. Kimura, "A Low-Impedance Open-Bitline Array for Multigigabit DRAM," *IEEE J. Solid-State Circuits* **37**, No. 4, 487–498 (April 2002).
37. S. Miyano and M. Takahashi, "Embedded DRAM SOCs and its Application for MPEG4 CODEC LSIs," *Proceedings of VLSI Circuits Short Course*, June 2001, pp. 101–121.
38. W. Leung, F.-C. Hsu, and M.-E. Jones, "The Ideal SoC Memory: 1T-SRAM," *Proceedings of the 13th Annual IEEE International ASIC/SOC Conference*, September 2000, pp. 32–36.
39. D. Somasekhar, S. Lu, B. Bloechel, K. Lai, S. Borkar, and V. De Planar, "1T-Cell DRAM with MOS Storage Capacitors in a 130nm Logic Technology for High Density Microprocessors Caches," *Proceedings of the ESSCIRC*, September 2002, pp. 127–130.
40. R. C. Foss, "Implementing Application Specific Memory," *ISSCC Digest of Technical Papers*, February 1996, pp. 260–261.
41. K. Nakazato, K. Itoh, H. Ahmed, H. Mizuta, T. Kisu, M. Kato, and T. Sakata, "Phase-State Low Electron-Number Drive Random Access Memory (PLEDM)," *ISSCC Digest of Technical Papers*, February 2000, pp. 132–133.
42. K. Noda, K. Matsui, K. Imai, K. Inoue, K. Tokashiki, H. Kawamoto, K. Yoshida, K. Takeda, N. Nakamura, T. Kimura, H. Toyoshima, Y. Koishikawa, S. Maruyama, T. Saitoh, and T. Tanigawa, "A 1.9- $\mu\text{m}^2$ /Loadless CMOS Four-Transistor SRAM Cell in a 0.18- $\mu\text{m}$  Logic Technology," *IEDM Tech. Digest*, pp. 643–646 (December 1998).
43. Y. Lin, C. Wu, C. Chang, R. Yang, W. Chen, J. Liaw, and C. Diaz, "Leakage Scaling in Deep Submicron CMOS for SoC," *IEEE Trans. Electron Devices* **49**, No. 6, 1034–1041 (June 2002).
44. Y. Ye, M. Khellah, D. Somasekhar, A. Farhang, and V. De, "A 6GHz, 16Kbytes L1 Cache in a 100nm Dual-V/Sub-T/ Technology Using a Bitline Leakage Reduction (BLR) Technique," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2002, pp. 50–51.
45. K. Ishibashi, K. Komiyaji, S. Morita, T. Aoto, S. Ikeda, K. Asayama, A. Koike, T. Yamanaka, N. Hashimoto, H. Iida, F. Kojima, K. Motohashi, and K. Sasaki, "A 12.5-ns 16-Mb CMOS SRAM with Common-Centroid-Geometry-Layout Sense Amplifiers," *IEEE J. Solid-State Circuits* **29**, No. 4, 411–418 (April 1994).
46. H. Sato, T. Wada, S. Ohbayashi, K. Kozaru, Y. Okamoto, Y. Higashide, T. Shimizu, Y. Maki, R. Morimoto, H. Otoi, T. Koga, H. Honda, M. Taniguchi, Y. Arita, and T. Shiomi, "A 500-MHz Pipelined Burst SRAM with Improved SER Immunity," *IEEE J. Solid-State Circuits* **34**, No. 11, 1571–1579 (November 1999).
47. K. Itoh, A. R. Fridi, A. Bellaouar, and M. I. Elmasry, "A Deep Sub-V, Single Power-Supply SRAM Cell with Multi-V<sub>t</sub>, Boosted Storage Node and Dynamic Load," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1996, pp. 132–133.
48. M. Yamaoka, K. Osada, and K. Ishibashi, "0.4-V Logic Library Friendly SRAM Array Using Rectangular-Diffusion Cell and Delta-Boosted-Array Voltage Scheme," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2002, pp. 170–173.
49. S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. Chandrakasan, "Scaling of Stack Effect and its Application for Leakage Reduction," *Proceedings of the ISLPED*, August 2001, pp. 195–199.
50. K. Seta, H. Hara, T. Kuroda, M. Kakumu, and T. Sakurai, "50% Active-Power Saving Without Speed



- Degradation Using Standby Power Reduction (SPR) Circuit," *ISSCC Digest of Technical Papers*, February 1995, pp. 318–319.
51. K. Kumagai, H. Iwaki, H. Yoshida, H. Suzuki, T. Yamada, and S. Kurosawa, "A Novel Powering-Down Scheme for Low Vt CMOS Circuits," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1998, pp. 44–45.
  52. M. Mizuno, K. Furuta, S. Narita, H. Abiko, I. Sakai, and M. Yamashina, "Elastic-Vt CMOS Circuits for Multiple On-Chip Power Control," *ISSCC Digest of Technical Papers*, February 1996, pp. 300–301.
  53. C. Akrouf, J. Bialas, M. Canada, D. Cawthron, J. Corr, B. Davari, R. Floyd, S. Geissler, R. Goldblatt, R. Houle, P. Kartschoke, D. Kramer, P. McCormick, N. Rohrer, G. Salem, R. Schulz, L. Su, and L. Whitney, "A 480-MHz RISC Microprocessor in a 0.12- $\mu\text{m}$   $L_{\text{eff}}$  CMOS Technology with Copper Interconnects," *IEEE J. Solid-State Circuits* **33**, No. 11, 1609–1616 (November 1998).
  54. H. Morimura and N. Shibata, "A 1-V 1-Mb SRAM for Portable Equipment," *Proceedings of the ISLPEP*, August 1996, pp. 61–66.
  55. K. Itoh, R. Hori, H. Masuda, Y. Kawajiri, H. Kawamoto, and H. Katto, "A Single 5V 64K Dynamic RAM," *ISSCC Digest of Technical Papers*, February 1980, pp. 228–229.
  56. I. Fukushi, R. Sasagawa, M. Hamaminato, T. Izawa, and S. Kawashima, "A Low-Power SRAM Using Improved Charge Transfer Sense Amplifiers and a Dual-Vth CMOS Circuit Scheme," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1998, pp. 142–145.
  57. M. Hasegawa, M. Nakamura, S. Narui, S. Ohkuma, Y. Kawase, H. Endoh, S. Miyatake, T. Akiba, K. Kawakita, M. Yoshida, S. Yamada, T. Sekiguchi, I. Asano, Y. Tadaki, R. Nagai, S. Miyaoka, K. Kajigaya, M. Horiguchi, and Y. Nakagome, "A 256Mb SDRAM with Subthreshold Leakage Current Suppression," *ISSCC Digest of Technical Papers*, February 1998, pp. 80–81.
  58. Y. Ye, S. Borkar, and V. De, "A New Technique for Standby Leakage Reduction in High-Performance Circuits," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1998, pp. 40–41.
  59. J. P. Halter and F. N. Najm, "A Gate-Level Leakage Power Reduction Method for Ultra-Low-Power CMOS Circuits," *Proceedings of the CICC*, May 1997, pp. 475–478.
  60. H. Kawaguchi, K. Nose, and T. Sakurai, "A Super Cut-Off CMOS (SCCMOS) Scheme for 0.5-V Supply Voltage with Picoampere Stand-By Current," *IEEE J. Solid-State Circuits* **35**, No. 10, 1498–1501 (October 2000).
  61. T. Yamagata, S. Tomishima, M. Tsukude, T. Tsuruda, Y. Hashizume, and K. Arimoto, "Low Voltage Circuit Design Techniques for Battery-Operated and/or Giga-Scale DRAM's," *IEEE J. Solid-State Circuits* **30**, No. 11, 1183–1188 (November 1995).
  62. T. Kuroda, T. Fujita, S. Mita, T. Nagamatu, S. Yoshioka, F. Sano, M. Norishima, M. Murota, M. Kako, M. Kinugawa, M. Kakumu, and T. Sakurai, "A 0.9V, 150MHz, 10mW, 4mm<sup>2</sup>, 2-D Discrete Cosine Transform Core Processor with Variable-Threshold-Voltage Scheme," *ISSCC Digest of Technical Papers*, February 1996, pp. 166–167.
  63. H. Mizuno, K. Ishibashi, T. Shimura, T. Hattori, S. Narita, K. Shiozawa, S. Ikeda, and K. Uchiyama, "A 18- $\mu\text{A}$  Standby Current 1.8-V 200-MHz Microprocessor with Self-Substrate-Biased Data-Retention Mode," *IEEE J. Solid-State Circuits* **34**, No. 11, 1492–1500 (November 1999).
  64. S. V. Kosonocky, M. Immediato, P. Cottrell, T. Hook, R. Mann, and J. Brown, "Enhanced Multi-Threshold (MTCMOS) Circuits Using Variable Well Bias," *Proceedings of the ISLPEP*, August 2001, pp. 165–169.
  65. H. Yamauchi, T. Iwata, H. Akamatsu, and A. Matsuzawa, "A 0.8V/100MHz/Sub-5mW-Operated Mega-Bit SRAM Cell Architecture with Charge-Recycle Offset-Source Driving (OSD) Scheme," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1996, pp. 126–127.
  66. A. Keshavarzi, S. Ma, S. Narendra, B. Bloechel, K. Mistry, T. Ghani, S. Borkar, and V. De, "Effectiveness of Reverse Body Bias for Leakage Control in Scaled Dual Vt CMOS ICs," *Proceedings of the ISLPEP*, August 2001, pp. 207–212.
  67. K.-S. Min, H. Kawaguchi, and T. Sakurai, "Zigzag Super Cut-Off CMOS (ZSCCMOS) Block Activation with Self-Adaptive Voltage Level Controller: An Alternative to Clock-Gating Scheme in Leakage Dominant Era," *ISSCC Digest of Technical Papers*, February 2003, pp. 400–401.
  68. S. Heo and K. Asanovic, "Leakage-Biased Domino Circuits for Dynamic Fine-Grain Leakage Reduction," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2002, pp. 316–319.
  69. K. Roy, "Leakage Sensitive Logic and Circuits," *Proceedings of the VLSI Circuits Short Course*, June 2002.
  70. G. Kitsukawa, M. Horiguchi, Y. Kawajiri, T. Kawahara, T. Akiba, Y. Kawase, T. Tachibana, T. Sakai, M. Aoki, S. Shukuri, K. Sagara, R. Nagai, Y. Ohji, N. Hasegawa, N. Yokoyama, T. Kisu, H. Yamashita, T. Kure, and T. Nishida, "256-Mb DRAM Circuit Technologies for File Applications," *IEEE J. Solid-State Circuits* **28**, No. 11, 1105–1113 (November 1993).
  71. T. Sakata, M. Horiguchi, M. Aoki, and K. Itoh, "Two-Dimensional Power-Line Selection Scheme for Low Subthreshold-Current Multi-Gigabit DRAMs," *Proceedings of the ESSCIRC*, September 1993, pp. 131–134.
  72. P. R. van der Meer and A. van Staveren, "New Standby-Current Reduction Technique for Deep Sub-Micron VLSI CMOS Circuits: Smart Series Switch," *Proceedings of the ESSCIRC*, September 2002, pp. 663–666.
  73. T. Kawahara, Y. Kawajiri, G. Kitsukawa, Y. Nakagome, K. Sagara, Y. Kawamoto, T. Akiba, S. Kato, Y. Kawase, and K. Itoh, "A Circuit Technology for Sub-10ns ECL 4Mb BiCMOS DRAMs," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 1991, pp. 131–132.
  74. H. Mizuno, N. Oodaira, Y. Kanno, T. Sakata, and T. Watanabe, "CMOS-Logic-Circuit-Compatible DRAM Circuit Designs for Wide-Voltage and Wide-Temperature-Range Applications," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2000, pp. 120–121.
  75. B. Wicht, J.-Y. Languier, and D. Schmitt-Landsiedel, "A 1.5V 1.7ns 4k $\times$ 32 SRAM with a Fully-Differential Auto-Power-Down Current Sense Amplifier," *ISSCC Digest of Technical Papers*, February 2003, pp. 462–463.
  76. Y. Kanno, H. Mizuno, N. Oodaira, Y. Yasu, and K. Yanagisawa, " $\mu\text{I/O}$  Architecture for 0.13- $\mu\text{m}$  Wide-Voltage-Range System-on-a-Package (SoP) Designs," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2002, pp. 168–169.
  77. W. Wen-Tai, K. Ming-Dou, C. Mi-Chang, and C. Chung-Hui, "Level Shifters for High-Speed 1 V to 3.3 V Interfaces in a 0.13 $\mu\text{m}$  Cu-Interconnection/Low-k CMOS Technology," *Proceedings of the International Symposium on VLSI Technology, Systems, and Applications*, May 2001, pp. 307–310.
  78. Y. Nakagome, K. Itoh, K. Takeuchi, E. Kume, H. Tanaka, M. Isoda, T. Musha, T. Kaga, T. Kisu, T. Nishida, Y. Kawamoto, and M. Aoki, "Circuit Techniques for 1.5–3.6-V Battery-Operated 64-Mb DRAM," *Proceedings of the ESSCIRC*, September 1990, pp. 157–160.

79. H. Sanchez, J. Siegel, C. Nicoletta, J. Alvarez, J. Nissen, and G. Gerosa, "A Versatile 3.3 V/2.5 V/1.8 V CMOS I/O Driver Built in a 0.2  $\mu\text{m}$  3.5 nm  $T_{\text{ox}}$  1.8 V CMOS Technology," *ISSCC Digest of Technical Papers*, February 1999, pp. 276–277.
80. G. P. Singh and R. B. Salem, "High-Voltage-Tolerant I/O Buffers with Low-Voltage CMOS Process," *IEEE J. Solid-State Circuits* **34**, No. 11, 1512–1525 (November 1999).
81. Y. Takemae, T. Ema, M. Nakano, F. Baba, T. Yabu, K. Miyasaki, and K. Shirai, "A 1Mb DRAM with 3-Dimensional Stacked Capacitor Cells," *ISSCC Digest of Technical Papers*, February 1985, pp. 250–251.
82. M. Kubo, R. Hori, O. Minato, and K. Sato, "A Threshold Voltage Controlling Circuit for Short Channel MOS Integrated Circuits," *ISSCC Digest of Technical Papers*, February 1976, pp. 54–55.
83. E. M. Blaser, W. M. Chu, and G. Sonoda, "Substrate and Load Gate Voltage Compensation," *ISSCC Digest of Technical Papers*, February 1976, pp. 56–57.
84. Y. Oowaki, M. Noguchi, S. Takagi, D. Takashima, M. Ono, Y. Matsunaga, K. Sunouchi, H. Kawaguchiya, S. Matsuda, M. Kamoshida, T. Fuse, S. Watanabe, A. Toriumi, S. Manabe, and A. Hojo, "A Sub-0.1 $\mu\text{m}$  Circuit Design with Substrate-Over-Biasing," *ISSCC Digest of Technical Papers*, February 1998, pp. 88–89.
85. M. Miyazaki, G. Ono, T. Hattori, K. Shiozawa, K. Uchiyama, and K. Ishibashi, "1000-MIPS/W Microprocessor Using Speed-Adaptive Threshold-Voltage CMOS with Forward Bias," *ISSCC Digest of Technical Papers*, February 2000, pp. 420–421.
86. K. Ishibashi, T. Yamashita, Y. Arima, I. Minematsu, and T. Fujimoto, "A 9 $\mu\text{W}$  50MHz 32b Adder Using a Self-Adjusted Forward Body Bias in SoCs," *ISSCC Digest of Technical Papers*, February 2003, pp. 116–117.
87. S. Kakimoto, T. Okuno, Y. Iwase, Y. Yaoi, F. Yoshioka, K. Kimoto, M. Nakano, K. Kawashima, S. Morishita, K. Sugimoto, T. Shiomi, T. Okumine, K. Kataoka, A. Shibata, S. Toyoyama, Y. Satoh, K. Fujimoto, K. Tatsumi, H. Kotaki, and A. Kito, "Self-Corrective Device and Architecture to Ensure LSI Operation at 0.5V Using Bulk Dynamic Threshold MOSFET with a Self-Adaptive Power Supply," *ISSCC Digest of Technical Papers*, February 2003, pp. 402–403.
88. J. Montanaro, R. T. Witek, K. Anne, A. J. Black, E. M. Cooper, D. W. Dobberpuhl, P. M. Donahue, J. Eno, A. Farrell, G. W. Hoepfner, D. Kruckemyer, T. H. Lee, P. Lin, L. Madden, D. Murray, M. Pearce, S. Santhanam, K. J. Snyder, R. Stephany, and S. C. Thierauf, "A 160 MHz 32 b 0.5 W CMOS RISC Microprocessor," *ISSCC Digest of Technical Papers*, February 1996, pp. 214–215.
89. D. R. Ditzel, "Transmeta's Crusoe: A Low-Power x86-Compatible Microprocessor Built with Software," presented at Cool Chips III, April 2000.
90. T. Burd, T. Pering, A. Stratakos, and R. Brodersen, "A Dynamic Voltage Scaled Microprocessor System," *ISSCC Digest of Technical Papers*, February 2000, pp. 294–295.
91. H. Mizuno and T. Kawahara, "ChipOS: Open Power-Management Platform to Overcome the Power Crisis in Future LSIs," *ISSCC Digest of Technical Papers*, February 2001, pp. 344–345.
92. T. Shimizu, F. Arakawa, and T. Kawahara, "Autonomous Decentralized Low-Power System LSI Using Self-Instructing Predictive Shutdown Method," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2001, pp. 55–56.
93. M. Miyazaki, G. Ono, H. Tanaka, N. Ohkubo, and T. Kawahara, "An Autonomous Decentralized Low-Power System with Adaptive-Universal Control for a Chip Multi-Processor," *ISSCC Digest of Technical Papers*, February 2003, pp. 108–109.
94. T. Miyake, T. Yamashita, N. Asari, H. Sekisaka, T. Sakai, K. Matsuura, A. Wakahara, H. Takahashi, T. Hiyama, K. Miyamoto, and K. Mori, "Design Methodology of High Performance Microprocessor Using Ultra-Low Threshold Voltage CMOS," *Proceedings of the CICC*, May 2001, pp. 275–278.
95. A. Bellaouar, A. Fridi, M. I. Elmasry, and K. Itoh, "Supply Voltage Scaling for Temperature-Insensitive CMOS Circuit Operation," *IEEE Trans. Circuits & Syst.* **45**, 415–417 (March 1998).
96. K. Kanda, K. Nose, H. Kawaguchi, and T. Sakurai, "Design Impact of Positive Temperature Dependence on Drain Current in Sub-1-V CMOS VLSIs," *IEEE J. Solid-State Circuits* **36**, No. 10, 1559–1564 (October 2001).
97. C. Chuang, P. Lu, and C. J. Anderson, "SOI for Digital CMOS LSI: Design Considerations and Advances," *Proc. IEEE* **86**, No. 4, 689–720 (April 1998).
98. J. M. Hill and J. Lachman, "A 900 MHz 2.25MB Cache with On-Chip CPU—Now in Cu SOI," *ISSCC Digest of Technical Papers*, February 2001, pp. 176–177.
99. R. V. Joshi, A. Pellela, O. Wagner, Y. H. Chan, W. Dachtera, S. Wilson, and S. P. Kowalczyk, "High Performance SRAMs in 1.5V, 0.18 $\mu\text{m}$  Partially Depleted SOI," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2002, pp. 74–77.
100. Y. Hirano, T. Iwamatsu, K. Shiga, K. Nii, K. Sonoda, T. Matsumoto, S. Maeda, Y. Yamaguchi, T. Ipposhi, S. Maegawa, and Y. Inoue, "High Soft-Error Tolerance Body-Tied SOI Technology," *Symposium on VLSI Technology, Digest of Technical Papers*, June 2002, pp. 48–49.
101. H. Sato, N. Itoh, K. Nii, K. Yoshida, Y. Nakase, H. Makino, A. Yamada, T. Arakawa, S. Iwade, Y. Hirano, and T. Ipposhi, "A 400MHz 183mW Microcontroller in Body-Tied SOI Technology," *ISSCC Digest of Technical Papers*, February 2003, pp. 110–111.
102. F. Morishita, K. Suma, M. Hirose, T. Tsuruda, Y. Yamaguchi, T. Eimori, T. Oashi, K. Arimoto, Y. Inoue, and T. Nishimura, "Leakage Mechanism Due to Floating Body and Countermeasure on Dynamic Retention Mode of SOI-DRAM," *Symposium on VLSI Technology, Digest of Technical Papers*, June 1995, pp. 141–142.
103. T. Yamada, K. Takahashi, H. Oyamatsu, H. Nagano, T. Sato, I. Mizushima, S. Nitta, T. Hojo, K. Kokubun, K. Yasumoto, Y. Matsubara, T. Yoshida, S. Yamada, Y. Tsunashima, Y. Saito, S. Nadahara, Y. Katsumata, M. Yoshimi, and H. Ishiuchi, "An Embedded DRAM Technology on SOI/Bulk Hybrid Substrate Formed with SEG Process for High-End SOC Application," *Symposium on VLSI Technology, Digest of Technical Papers*, June 2002, pp. 112–113.
104. F. Assaderaghi, S. Parke, P. K. Ko, and C. Hu, "A Novel Silicon-On-Insulator (SOI) MOSFET for Ultra Low Voltage Operation," *Proceedings of the ISLPED*, August 1994, pp. 58–59.
105. T. Douseki, J. Yamada, and H. Kyuragi, "Ultralow-Power CMOS/SOI LSI Design for Future Mobile Systems," *Symposium on VLSI Circuits, Digest of Technical Papers*, June 2001, pp. 6–9.
106. H. Kawaguchi, K. Kanda, K. Nose, S. Hattori, D. D. Antono, D. Yamada, T. Miyazaki, K. Inagaki, T. Hiramoto, and T. Sakurai, "A 0.5V, 400MHz,  $V_{\text{DD}}$ -Hopping Processor with Zero- $V_{\text{TH}}$  FD-SOI Technology," *ISSCC Digest of Technical Papers*, February 2003, pp. 106–107.
107. T. Douseki, T. Shimamura, and N. Shibata, "A 0.3V 3.6GHz 0.3mW Frequency Divider with Differential ED-CMOS/SOI Circuit Technology," *ISSCC Digest of Technical Papers*, February 2003, pp. 114–115.
108. K. Takeuchi, R. Koh, and T. Mogami, "A Study of the Threshold Voltage Variation for Ultra-Small Bulk and

- SOI CMOS," *IEEE Trans. Electron Devices* **48**, No. 9, 1995–2001 (September 2001).
109. D. Hisamoto, "FD/DG-SOI MOSFET: A Viable Approach to Overcoming the Device Scaling Limit," *IEDM Tech. Digest*, pp. 429–432 (December 2001).
  110. J. Kedzierski, E. Nowak, T. Kanarsky, Y. Zhang, D. Boyd, R. Carruthers, C. Cabral, R. Amos, C. Lavoie, R. Roy, J. Newbury, E. Sullivan, J. Benedict, P. Saunders, K. Wong, D. Kanaperi, M. Krishnan, K.-L. Lee, B. A. Rainey, D. Fried, P. Cottrell, H.-S. P. Wong, M. Jeong, and W. Haensch, "Metal-Gate FinFET and Fully-Depleted SOI Devices Using Total Gate Silicidation," *IEDM Tech. Digest*, pp. 247–250 (December 2002).
  111. C. P. Auth and J. D. Plummer, "Scaling Theory for Cylindrical, Fully Depleted Surrounding Gate MOSFETs," *IEEE Electron Device Lett.* **18**, 74–76 (February 1997).
  112. F. H. Gaensslen and R. C. Jaeger, "Low Temperature Microelectronics," *Extended Abstracts of the International Conference on Solid-State Devices and Materials*, August 1990, pp. 353–356.
  113. P. K. Naji, M. Durlam, S. Tehrani, J. Calder, and M. F. DeHerrera, "A 256kb 3.0V 1T1MTJ Nonvolatile Magnetoresistive RAM," *ISSCC Digest of Technical Papers*, February 2001, pp. 122–123.
  114. M. Gill, T. Lowrey, and J. Park, "Ovonic Unified Memory: A High-Performance Nonvolatile Memory Technology for Stand-Alone Memory and Embedded Applications," *ISSCC Digest of Technical Papers*, February 2002, pp. 202–203.

Received November 21, 2002; accepted for publication March 24, 2003

**Yoshinobu Nakagome** *Renesas Technology Corporation, Kodaira, Tokyo, 187-8588, Japan (nakagome.yoshinobu@renesas.com)*. Mr. Nakagome received his B.S. degree in electrical and electronic engineering and his M.S. degree in applied electronics from the Tokyo Institute of Technology in 1978 and 1980, respectively. In 1980, he joined the Central Research Laboratory, Hitachi Ltd., Tokyo, where he was engaged in research on MOS device physics and technologies. From 1983 to 1987, he worked on high-density MOS dynamic memories. He was a visiting Industrial Fellow at the University of California, Berkeley, from 1987 to 1988. From 1988 to 1992, he led the research project on 64Mb DRAM featuring 1.5-V operation. From 1992 to 1995, he was a manager of the research group of embedded DRAM, FeRAM, and 1Gb DRAM. From 1995 to 1999, he was with the Semiconductor Technology Development Center, Semiconductor & Integrated Circuits, Hitachi Ltd., where he led a product development team of 256-Mb SDRAM and DDR SDRAM. In 1999 he assumed his position as the department manager covering low-power technologies for memory LSIs and system LSIs, and nonvolatile memory technologies. From 2001 to 2003, he was the department manager of the Analog in System Development Department, Semiconductor & Integrated Circuits, Hitachi Ltd. Since April 2003, he has been with the Renesas Technology Corporation. Mr. Nakagome has authored or co-authored 40 technical publications and holds more than 40 patents in Japan and more than 30 patents in the U.S. He was one of the recipients of the Best Paper Award of ESSCIRC'90. Mr. Nakagome was an associate editor for the *IEEE Journal of Solid-State Circuits* from 1996 through 1999. He was a technical program committee member of the Symposium on VLSI Circuits from 1994 through 1997. He was a program co-chairman and chairman of the 2002 and 2003 Symposium on VLSI Circuits, respectively. Mr. Nakagome is a member of the Institute of Electronics, Information, and Communication Engineers of Japan.

**Masashi Horiguchi** *Renesas Technology Corporation, Kodaira, Tokyo, 187-8588, Japan (horiguchi.masashi@renesas.com)*. Dr. Horiguchi received his B.S. degree in electronic engineering in 1977, his M.S. degree in information engineering in 1979, and a Ph.D. degree in electrical engineering from the University of Tokyo in 2000. He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, in 1979. He has been engaged in the research and development of MOS dynamic-memory circuit technologies, including on-chip voltage converters, redundancy circuits, and subthreshold-leakage current reduction techniques. In 1995, he moved to Semiconductor & Integrated Circuits, Hitachi Ltd., where he worked on the development of DRAMs, pseudo SRAMs, and on-chip voltage converter circuits. Since 2003 he has been with the Renesas Technology Corporation. Dr. Horiguchi is a member of the IEEE and the Institute of Electronics, Information, and Communication Engineers of Japan. He has authored or co-authored 20 technical publications and holds more than 50 patents in Japan and more than 70 patents in the U.S.

**Takayuki Kawahara** *Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo, 185-8601, Japan (tkawaha@crl.hitachi.co.jp)*. Dr. Kawahara received his B.S. and M.S. degrees in physics in 1983 and 1985, respectively, and his Ph.D. degree in electronic engineering in 1993 from Kyusyu University, Japan. In 1985, he joined the Central Research Laboratory, Hitachi Ltd., Tokyo. Since then, he has made fundamental contributions in many areas in the field of low-

power high-speed memories. In the field of DRAM circuits from 1985 to 1993, his major contributions concerned low-power, low-voltage circuits including subthreshold-current reduction by gate-source self-reverse biasing technique and an overdrive sense-amplifier scheme coupled with direct sensing. He also pioneered the charge-recycling scheme, a concept now widely applied to various circuits. In the field of flash memory, from 1994 to 2001, he and his team developed a bit-line clamped sensing scheme for fast sensing, a high-voltage generator scheme under a low voltage supply, and a pioneering high-speed programming method. Currently, he is leading research groups of DRAMs, SRAMs, and nonvolatile memories in the laboratory. He has published more than 30 papers in the *IEEE Journal* and at IEEE-sponsored conferences. He holds 58 U.S. and Japanese patents. He was a visiting researcher at Electronics Laboratory (LEG), Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland, from 1997 to 1998. Dr. Kawahara has been a member of the ISSCC program committee in Memory, Technology Direction, and Far East since 2000.

**Kiyoo Itoh** *Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo, 185-8601, Japan (k-itoh@crl.hitachi.co.jp).*

Dr. Itoh received his B.S. and Ph.D. degrees in electrical engineering from Tohoku University, Japan, in 1963 and 1976, respectively. He is currently one of three Fellows at Hitachi Ltd. He was a visiting MacKay Lecturer at the University of California, Berkeley, in 1994, a visiting professor at the University of Waterloo in 1995, and a consulting professor at Stanford University from 2000 to 2001. Since 1972, he has led DRAM circuit technology at Hitachi Ltd. He was the lead designer of the first prototype for eight generations of Hitachi DRAMs ranging from 4Kb to 64Mb. As early as 1988, as a pioneer, he also developed low-power/low-voltage CMOS circuits focusing on subthreshold current reduction. He holds more than 180 patents in Japan, and more than 140 patents in the U.S., including the folded bit line. He has authored and co-authored three books on memory designs, and more than 120 papers in IEEE journals and conference proceedings. Dr. Itoh has won many honors, including the IEEE Paul Rappaport Award in 1984, the Best Paper Award of ESSCIRC'90, and the 1993 IEEE Solid-State Circuits Award. He is an IEEE Fellow. In Japan, Dr. Itoh's awards include the Commendation by the Minister of State for Science and Technology (Person of Scientific and Technological Merits) in 1997, and a National Medal of Honor with Purple Ribbon in 2000.