Security strategy of powered-off SRAM for resisting physical attack to data remanence*

Yu Kai(余凯), Zou Xuecheng(邹雪城)†, Yu Guoyi(余国义), and Wang Weixu(王伟旭)

(Department of Electronic Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: This paper presents a security strategy for resisting a physical attack utilizing data remanence in powered-off static random access memory (SRAM). Based on the mechanism of physical attack to data remanence, the strategy intends to erase data remanence in memory cells once the power supply is removed, which disturbs attackers trying to steal the right information. Novel on-chip secure circuits including secure power supply and erase transistor are integrated into conventional SRAM to realize erase operation. Implemented in 0.25 μm Huahong-NEC CMOS technology, an SRAM exploiting the proposed security strategy shows the erase operation is accomplished within 0.2 μs and data remanence is successfully eliminated. Compared with conventional SRAM, the retentive time of data remanence is reduced by 82% while the operation power consumption only increases by 7%.

Key words: SRAM; security strategy; physical attack; data remanence; low-voltage low-power

DOI: 10.1088/1674-4926/30/9/095010 **EEACC:** 2570

1. Introduction

Static random access memory (SRAM) is widely used in many security systems and smartcards for temporary storage of secret data or crypto keys. It is commonly believed that the information represented by data from such memory disappears immediately after the power supply is removed. However, data remanence caused by some physical effects in powered-off SRAM is reported[1−3], and many experiments show the retentive time of data remanence is increased at low temperatures[3,4]. As a result, the possibility of a novel physical attack to data remanence in the powered-off SRAM is rising[5]. Meanwhile, the study of security strategy for SRAM to resist this kind of physical attack has become a hot topic[6].

The conventional security strategy for SRAM utilizes tamper-sensing enclosure embedded into memory. On detection of a tampering event, the memory is powered off, which is regarded as a successful protection. Obviously, this strategy fails to resist a physical attack to data remanence in powered-off SRAM. If the retentive time of data remanence exceeds the time required by attackers to power up the memory again, especially at low temperature conditions to freeze data remanence, the data will be successfully recovered. Therefore, this kind of protection mechanism will be defeated.

In this article, we present a security strategy to resist a physical attack to data remanence in powered-off SRAM. This strategy uses novel secure circuits including a secure power supply and erase transistor to eliminate data remanence in memory cells once the power supply is removed. Moreover, all secure circuits are fully integrated into conventional SRAM to avoid being disassembled by attackers.

In addition, the mechanism of the physical attack to data remanence is analyzed, which is the theoretical basis for designing a security strategy. The system architecture of SRAM exploiting the security strategy is outlined, and the implementation of secure circuits is described

2. Mechanism of physical attack to data remanence

Data remanence in powered-off SRAM originates from various physical effects, such as carrier transport processes, ionic contamination, and hot carriers[1−3]. However, the carrier transport process is proved to be the main factor leading to data remanence[3], which is mostly analyzed as follows. Figure 1 presents the structure of a typical six-transistor cell in SRAM. Two inverters are built from pairs of NFETs (M1, M2) and PFETs (M3, M4). The output of one inverter is coupled to the input of the other, and vice versa. Two other NFETs (M5, M6) are used to control reading data from and writing data into.

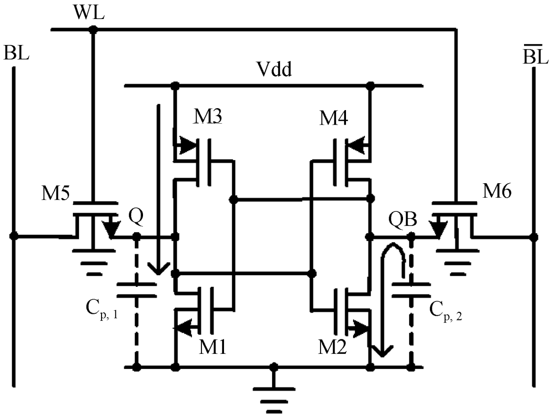

Fig. 1. Schematic of a six-transistor cell in SRAM.

* Project supported by the National Natural Science Foundation of China (No. 60776027).

† Corresponding author. Email: estxczou@gmail.com

Received 18 February 2009, revised manuscript received 7 April 2009

We assume data ‘1’ is stored in this cell. Then M1 and M4 are cut off while M2 and M3 are turned on. The power supply charges parasitic capacitance $C_{\mathrm{p},1}$ while M2 discharges parasitic capacitance $C_{\mathrm{p},2}$. As M2 is in the triode region, electrons accumulate in the surface of Si–SiO$_2$ to form an inversion layer. Meanwhile, a depletion region consisting of fixed negative space charge is under the inversion layer. The electrons and negative space charge are induced to shield electric fields generated by the gate potential of M2. When the power supply is removed, external electric fields suddenly disappear. The electrons enriched near the surface only depend on diffusion to recombine with holes in the p-type body. Moreover, the depletion region will disappear until the holes diffusing from the p-type body to the surface neutralize negative space charge. In this process, positive charge is induced by electrons and negative space charge and reserved in $C_{\mathrm{p},1}$, which is in accordance with the storage of data ‘1’ and leads to data remanence. As long as the whole p-type body including the Si–SiO$_2$ surface reaches a new dynamic equilibrium and concentrations of carriers (electrons and holes) are equal everywhere, data remanence will vanish.

The mechanism of physical attack to data remanence is based on the fact that diffusion and recombination rates of carriers (electrons and holes) are both in inverse proportion to the square root of temperature[2, 3]. So the induced positive charge in $C_{\mathrm{p},1}$ can hold from seconds to hours at low temperatures[4, 5]. Attackers just need to freeze powered-off SRAM by a cooling spray or liquid nitrogen and then power it up again. If the time required by the attackers is within the retentive time of data remanence, the previous data ‘1’ will be recovered. As each memory cell is symmetrical, the analysis on storage of data ‘0’ is the same as on storage of data ‘1’.

3. System architecture of security SRAM

According to the previous analysis of the mechanism of physical attack to data remanence, we propose a security strategy based at circuit and system level, which is more convenient and lower in cost compared with process and device solutions. The strategy intends to erase data remanence in memory cells when the power supply is removed, which reduces the retentive time of data remanence and disturbs attackers trying to steal the right information. Figure 2 shows the system architecture of SRAM employing the proposed security strategy, which is defined by us as security SRAM to distinguish it from conventional SRAM. The whole system includes an SRAM core with erase transistor and secure power supply.

The SRAM core is composed of a memory cell array, row/column decoders, and sense amplifiers/drivers. Address lines (A_1–A_m) select a particular row while other address lines (A_1–A_n) select a particular column. Then the cell of interest is accessed for reading data from or writing data into. The erase transistor is directly integrated into the memory cell array to eliminate data remanence without the operation of decoders and drivers, which is beneficial for achieving lower power consumption.

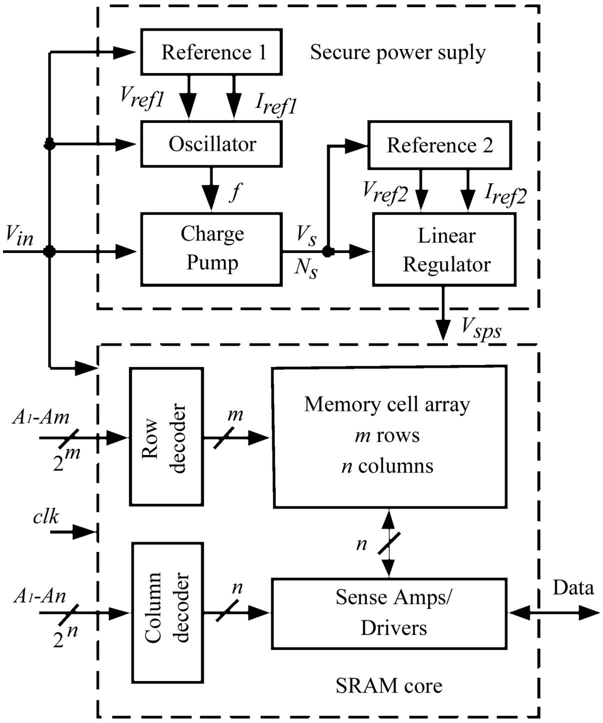

Fig. 2. System architecture of security SRAM.

The secure power supply which integrates charge pump, linear regulator, oscillator, and two references has two operation modes in terms of power supply. In powered-on mode, the secure power supply absorbs energy from the power supply and stores it into an on-chip storage device, while in powered-off mode, it releases the reserved energy to the memory cell array to accomplish erase operation.

4. Implementation of secure circuits

4.1. SRAM core with erase transistor

Compared to the conventional structure, erase transistor is embedded into each memory cell of the SRAM core, as shown in Fig. 3. We select NFET to serve as the erase transistor. As long as the power supply is removed, the erase transistor is selected by the secure power supply voltage V_{sps}, which establishes an inversion layer in it. As a result, the induced charge reserved in parasitic capacitances is neutralized through this erase transistor. The data remanence in memory cells is eliminated after the power-off occurs.

4.2. Secure power supply

The secure power supply is the key block in the whole system. The first design essential for it is the selection of a storage device which is connected between node N_{s} and ground. We adopt a high-voltage on-chip capacitor C_{s}, shown in Fig. 4, instead of an inductor which has larger series resistance, consuming more power. Moreover, an inductor produces electromagnetic interference, which induces currents in nearby conductors and substrate.

The second design essential for secure power supply is to enlarge the reserved energy W_{s} in the storage capacitor for low

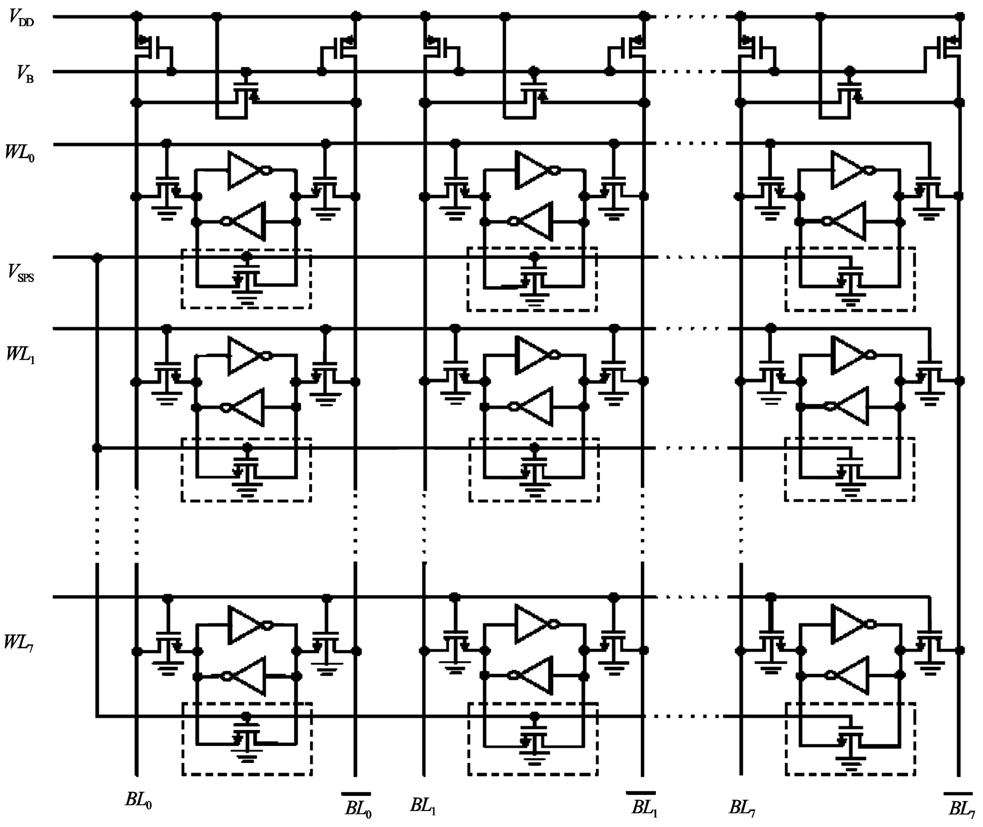

Fig. 3. Memory cell array with erase transistor in the SRAM core.

voltage application. Compared to improving the capacity C_s, it is more efficient to enhance the voltage V_s of the storage capacitor, which is determined by:

$$W_s = \frac{1}{2} C_s V_s^2. \tag{1}$$

So a charge pump is used in the secure power supply to elevate the power supply voltage V_{in} to a higher voltage V_s with multiple coefficient N. Meanwhile, the charge pump serves as a back-biased diode in powered-off mode to prevent energy in the storage capacitor reversely discharging to the power supply.

The topology of the charge pump, which includes a two-stage voltage doubler[6], is shown in Fig. 4. Each stage is controlled by two phase complementary clocks clk_1 and clk_2 having 50% duty cycle which is generated by an oscillator. Clock signals are coupled to gates of switches Mn,i and Mp,i through a pump capacitor C_i, thus causing a charge packet to flow through Mp,1 from capacitor C_1 to the first stage output N_h and, hence, to capacitor C_4 when clk_1 is high (clk_2 low) while, when clk_1 is low (clk_2 high), a charge packet flows through Mn,1 from V_{in} (stage input) into capacitor C_1, thus restoring the charge lost during the preceding phase. So packets of charge from the power supply are transferred into the storage capacitor through the charge pump. As we use the low-cost P-sub N-well technology, the body of the NFET does not short-circuit to its source.

The third design essential is to convert the reserved energy into voltage and current more efficiently. In powered-off mode, the equations for voltage V_s and current I_s with time are written as:

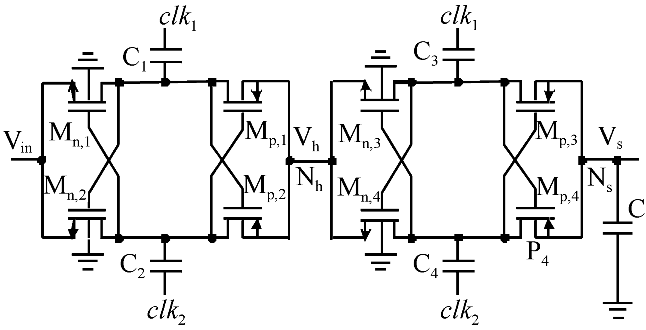

Fig. 4. Schematic of the charge pump with storage capacitor.

$$V_s = N V_{in} \exp\left(-\frac{t}{R_l C_s}\right), \tag{2}$$

$$I_s = -\frac{N V_{in}}{R_l} \exp\left(-\frac{t}{R_l C_s}\right), \tag{3}$$

where R_l is load resistance. So the linear regulator shown in Fig. 5 is used in the secure power supply to modulate the exponential voltage V_s into a constant standard voltage V_{sps} ($V_{sps} = V_{in}$) and transfer exponential current I_s efficiently in erase operation.

To reduce power consumption, the sampling network adopts PFETs fabricated in N-wells with the same threshold voltage. Their aspect ratio is designed to be $(W/L)_{18}/(W/L)_{19} = 1$ to obtain exact sampling. Moreover, M13–M16 are used as powered-off switches to the operation of the control linear regulator. In powered-on mode, M14 is turned on by the power supply voltage V_{in}, which results in M16 being in the triode

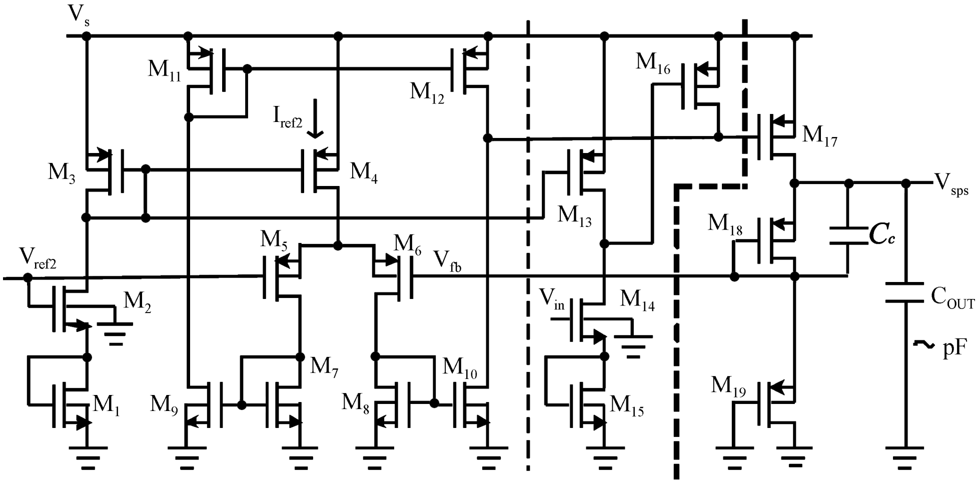

Fig. 5. Schematic of the linear regulator with on-chip output capacitor.

region and the pass transistor M17 is cut off; in powered-off mode, M14 and M16 are cut off while the pass transistor M17 is in operation, which allows the linear regulator to transfer the secure power supply to the SRAM core.

In addition, the secure power supply also integrates the oscillator and two references. The oscillator, based on *RC* multivibrator topology, produces two phase non-overlapping clock signals to drive the charge pump. The two references generate a reference current and voltage to the oscillator and linear regulator. Reference [1] is designed for low-voltage operation and obtains process, voltage, and temperature independent sources. However, Reference [2] is designed for high-voltage operation and required ultra-low power consumption.

5. Experimental results

To verify the proposed strategy, an 8 × 8 bit 6-T security SRAM integrating secure circuits and an 8 × 8 bit 6-T conventional SRAM have been designed with 0.25 μm Huahong-NEC CMOS technology. The power supply is 2.5 V while the operation frequency is 20 MHz. We assume the powered-down process continues for 1 μs with constant slope of 2.5 V/μs, which is a typical value in most SRAMs.

Figure 6 shows the power-up and power-down processes of a security SRAM. From the figure, in the power-up process, the voltage of the storage capacitor V_s follows the power supply voltage V_{in} with a delay (~3 μs) due to the setup of the charge pump while the secure power supply voltage V_{sps} remains at a low level. In the power-down process, V_{in} drops to ground quickly, which results in V_s having an equal voltage drop. Then V_s decreases slowly and exponentially with time and the secure power supply outputs the regulated voltage V_{sps}. So, the role of the secure power supply is verified.

Q/QB represent data stored in some memory cell of the conventional SRAM while Q'/QB' represent data stored in the same cell in the security SRAM. From Fig. 7, the original data in two memory cells are both data '1'. After the power dump occurs, Q decreases slowly and QB maintains a low level, which means the induced positive charge is held in the parasitic capacitance $C_{p,1}$ and data remanence is not eliminated. However, Q'/QB' becomes equal quickly, within 0.2 μs, when the power supply is removed, which means the induced positive charge is shared between the two parasitic capacitances and data remanence is eliminated.

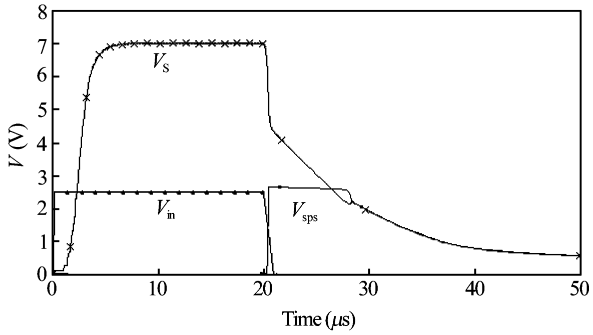

Fig. 6. Power-up and power-down processes of a security SRAM.

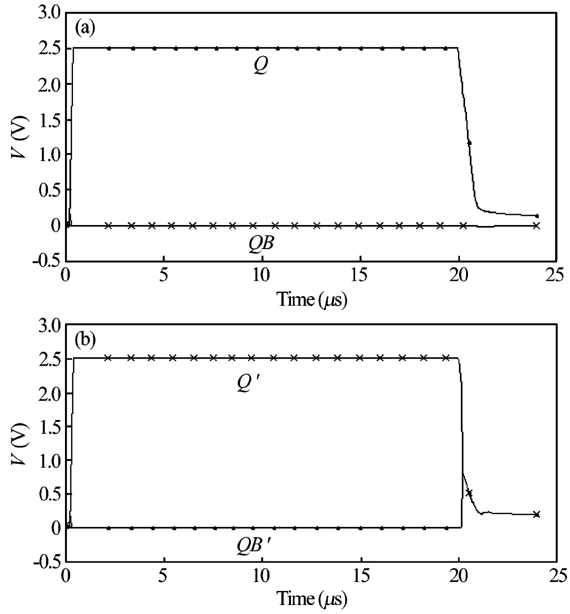

Fig. 7. Data remanence in the same cell of the security SRAM and a conventional SRAM: (a) Without erase transistor; (b) With erase transistor.

The transition currents of the erase transistor are shown

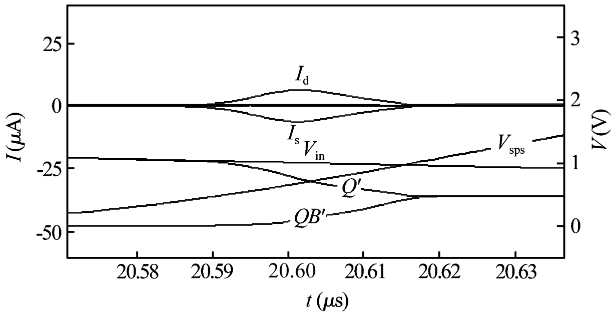

Fig. 8. Transition currents of the erase transistor in the power-down process.

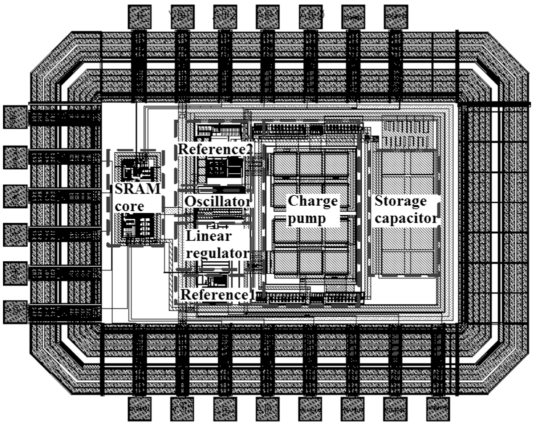

Fig. 9. Layout of the security SRAM.

in Fig. 8. In the powered-on mode, four terminal currents of the erase transistor are zero. Then in the power-down process, the source and drain currents of the erase transistor are the same with opposite direction, which signifies that the induced charge is neutralized. Meanwhile, the substrate and gate currents are approximately zero. So, the correct operation of the erase transistor is also verified. The strategy successfully erases data remanence in memory cells when the power supply is removed, which reduces the retentive time of data remanence and disturbs attackers trying to steal the right information.

Figure 9 shows the layout of the whole system which occupies an area of 850 × 1150 μm^2. The 8 × 8 bit 6-T SRAM core just occupies an area of 300 × 500 μm^2. It seems that the introduction of secure circuits takes up most of the silicon area. However, the secure power supply integrates a 100 pF storage capacitor and generates a 300 μA current while each cell only consumes 60 nA to drive the erase transistor, which means the largest storage capacity is 5 kbit. That is to say, if we design the storage capacity of the SRAM core to be 2 kbit, the silicon area percentage of secure circuits in the whole system will be reduced to 13%. This capacity is sufficient to keep secret data or crypto keys.

The arrangement of the layout should be carefully considered to reduce interference. The SRAM core, which is the most sensitive part, is placed far away from the storage capacitor. Meanwhile, different blocks are isolated by the safe-guard rings. Moreover, for the PAD arrangement, the side near the storage capacitor is deserted to arrange PAD and avoid signal interference.

Table 1. Comparison of characteristics between conventional SRAM and security SRAM.

Characteristics	Conventional SRAM	Security SRAM
Read access time (ns)	12	12
Write access time (ns)	10	10
Static current (μA)	5	5
Operation power consumption (μW)	1225	1325
Retentive time of data remanence (μs)	10	1.6

Table 1 summaries the comparison of characteristics between conventional SRAM and security SRAM based on the experimental results. The read and write access times of the two SRAMs are the same, which proves that the introduction of the erase transistor has little effect on the parasitic capacitances of each memory cell in normal operation. Meanwhile, secure circuits only consume 125 μW which causes the operation power consumption of the memory to only increase by 7%. Compared with conventional SRAM, the retentive time of data remanence is reduced by 82%.

6. Conclusions

This paper presents a security strategy to resist a physical attack to SRAM when the power supply has been removed. Secure circuits including a secure power supply and erase transistor are proposed. An SRAM integrated secure circuit and a conventional SRAM are designed with 0.25 μm Huahong-NEC CMOS technology. Experimental results validate the proposed security strategy, confirming that it successfully eliminates data remanence and reduces retentive time, which disturbs attackers trying to steal the right information. Moreover, compared with conventional SRAM, the operation power consumption only increases by 7% while the read and write operations of the two SRAMs are the same.

References

[1] Gutmann P. Secure deletion of data from magnetic and solid-state memory. 6th USENIX Security Symposium Proceedings, San Jose, California, 1996

[2] Skorobogatov S. Low temperature data remanence in static RAM. Technology Report of University of Cambridge, 2002

[3] Jiao H F, Zhang X B, Jia X Z, et al. The characteristic study of data remanence of SRAM. Research & Progress of SSE, 2006, 26(4): 536

[4] Samyde D, Skorobogatov S, Anderson R, et al. On a new way to read data from memory. IEEE Computer Society, 2003

[5] Guttmann P. Data remanence in semiconductor devices. The 10th Usenix Security Symposium, 2001: 1

[6] Pan F, Samaddar T. Charge pump circuit design. New York: McGraw-Hill, 2006: 169

[7] Milliken R J, Silva-Martínez J, Sánchez-Sinencio E. Full on-chip CMOS low-dropout voltage regulator. IEEE Trans Circuits Syst I, 2007, 54(9): 1879